Blind Source Separation : The Effects of Signal Non-Stationarity

Marcus J. T. Alphey

A thesis submitted for the degree of Doctor of Philosophy.
The University of Edinburgh.
March 2002

Abstract

This thesis investigates the effect of non-stationarity reduction, in the form of silence removal, on the performance of blind separation and deconvolution techniques for speech signals. An information-maximisation-based system is used for the separation of instantaneously mixed signals, and a decorrelating system for convolutively mixed signals.

An introduction to the concepts of adaptive signal processing, blind signal processing and artificial neural networks is presented. A review of approaches to solving the blind signal separation and deconvolution problems is provided. The susceptibility of the information-maximisation approach to signal non-stationarity is discussed and two methods of silence identification and removal are compared and used to pre-process data before blind separation.

The "infomax" approach is used to separate instantaneous mixtures, and is also modified to incorporate silence assessment and removal techniques to form an on-line system. Further modifications are made to the algorithm to investigate the effect of alternative update strategies, and these are compared with experimental results from identical modifications to diverse separating algorithms. A performance metric is used to assess the quality of separation achieved.

The application of these techniques to convolutively mixed speech signals is also investigated, using the CoBliSS algorithm. The effectiveness of the application of the silence removal techniques to both the time domain and frequency domain representations of the outputs is tested.

While this form of non-stationarity reduction improves the rate of convergence for instantaneous mixtures, it does not cause any significant improvement in separation performance under most of the experimental conditions tested. No significant difference in performance was noted for the separation of convolutive mixtures in either the time or frequency domain.

Acknowledgements

Of the many people to whom I owe a great deal of thanks for their efforts in helping me through my PhD, the following deserve special mention :

- Professor Alan Murray and Dr. David Laurenson, my academic supervisors here in the department, for their advice, guidance and encouragement throughout my period of study.
- Professor Bernie Mulgrew and Dr. Alister Hamilton for their advice and support during the latter part of my research.
- GEC Marconi Avionics (now BAE SYSTEMS Avionics Ltd) — in particular Dr. Paul Holbourn, my industrial supervisor, for funding my research and the other staff for their support and encouragement.
- Mr Jim McNicol, Principal Statistician at BioSS, for advice on the application of statistics and interpretation of results.
- BT Labs, Martlesham Heath for some of the data used in the experiments.
- David Stewart and the rest of the E & EE Departmental Computing Support team for all their efforts and assistance over the years.
- My parents and my brother and sister, for always being there for me, putting up with me through it all and for their continual encouragement.
- My flatmates Graham Clark, George Taylor and all those since, for our once high-tech flat's near legendary status, and for all the good times.
- Catherine Breslin for helping me see a great many things from a different perspective.
- Ashley McIntyre for her encouragement, support, patience and love.

From the ISG, I must single out Robin Woodburn and Emma Braithwaite for repeatedly assuring me that it could be done. Thanks also go to Pete Edwards for valiantly leading the coffee-time herd, Neil Marston, Andy Connelly and Mark Glover for finishing so far ahead of me, Ryan Dalzell for all those technical discussions, and on why we were actually here in the

first place, and Aslı Arslan for showing what can be achieved in a year. Finally, I would like to mention Ben Hounsell, Mark Harding, Alasdair Sutherland, Andrew Peacock, Patrice Fleury and Adria Bofill for making the lab bearable during my rework and write-up.

To the rest of the group, and the other occupants of Rooms 2.3, 2.2, 2.7 and 2.6 — those who were there before I arrived, as well as those who have come and gone since — I offer my thanks for the great coffee- and lunch-time banter, and the atmosphere in the lab.

Contents

List of figures

List of tables

List of acronyms and abbreviations

ANOVA	Analysis of Variance
ANN	Artificial Neural Network
ATM	Asynchronous Transfer Mode
aVLSI	analogue VLSI
BSS	Blind Signal (or Source) Separation
CDF	Cumulative Density Function
CDMA	Code Division Multiple Access
CoBliSS	Convolutive Blind Signal Separation
CV	Co-efficient of Variation
DAT	Digital Audio Tape
DOA	Direction of Approach
DSP	Digital Signal Processor
EASI	Equivariant Adaptive Separation via Independence
EEG	Electro-encephalographic
ELR	Energy-based Learning Rate (see Section 5.7.1)
EPP	Exploratory Projection Pursuit
FFT	Fast Fourier Transform
FIR	Finite Impulse Response
FPGA	Field Programmable Gate Array
HMM	Hidden Markov Model
HOS	Higher Order Statistics
ICA	Independent Component Analysis
IFFT	Inverse Fast Fourier Transform
IIR	Infinite Impulse Response
ISI	Inter-Symbol Interference
JADE	Joint Approximate Diagonalization of Eigen-Matrices
LSD	Least Significant Difference
ML	Maximum Likelihood
MLE	Maximum Likelihood Estimation
OFDM	Orthogonal Frequency Division Multiplexing
PCA	Principal Component Analysis
PDF	Probability Density Function
SED	Standard Error of Difference
SIR	Signal-to-Interference Ratio
VLSI	Very Large Scale Integration
VTOA	Voice Telephony Over ATM

Chapter 1
Introduction

This chapter gives a general introduction to the work presented in this thesis, providing an overview of the research field, indicating current areas of interest and identifying the focus of the investigation. The aims of this work are then addressed, followed by an outline of the content of the other chapters.

1.1 Blind signal processing

Blind signal processing is an exciting area of research, currently enjoying a great deal of interest from a wide range of fields including neural networks, signal processing and mathematics. Collaboration between these different communities has lead to improved understanding of the problem bounds, and has produced several useful advances in solution strategies.

So-called *blind* signal processing problems form a specific class of problems in which there is no *a priori* knowledge about the signals being processed. Consequently, the problems require a special set of solutions, namely those that can be extended or generalised to deal with as wide a range of instances as possible of the particular problem being addressed, regardless of situation-specific factors such as the type or number of signals being dealt with. Many common solutions to non-blind problems rely on simplifications that make use of key characteristics of the signals involved or of restrictions on the problem configuration. Solutions to blind problems cannot afford this reliance. The techniques used to solve blind problems must be more robust and reliably produce valid solutions without any assumptions about the input signals, or their environment, other than those implicit in the problem specification.

Such methods are therefore useful in investigative or analytical work where nothing is known in advance about the observed signals, and in real-world applications where characteristics of the signals and the environment may be subject to change. In such cases, the methods used must also be adaptive, allowing them to modify their behaviour to compensate for changes in the specific problem configuration.

1.1.1 Blind signal separation

An example of a blind problem is that of blind signal (or source) separation (BSS). The basic problem referred to here is to separate m observed linear combinations of n independent signals back into their respective sources, without any *a priori* knowledge of these sources, or their mixing. In the simplest case, it is assumed that there are no temporal complications, such as echoes or multi-path signals, and that the mixing is therefore an instantaneous process. The more complicated cases that do take account of temporal delays and filtering effects are usually referred to as the blind deconvolution problem. Furthermore, these mixing processes must be reversible, otherwise no solution is possible. Finally, if a complete separation is to be achieved, there must be at least as many observed mixtures as there are sources ($m \geq n$). The situation is illustrated by Figure 1.1. Finding a solution to this problem requires solving a system of

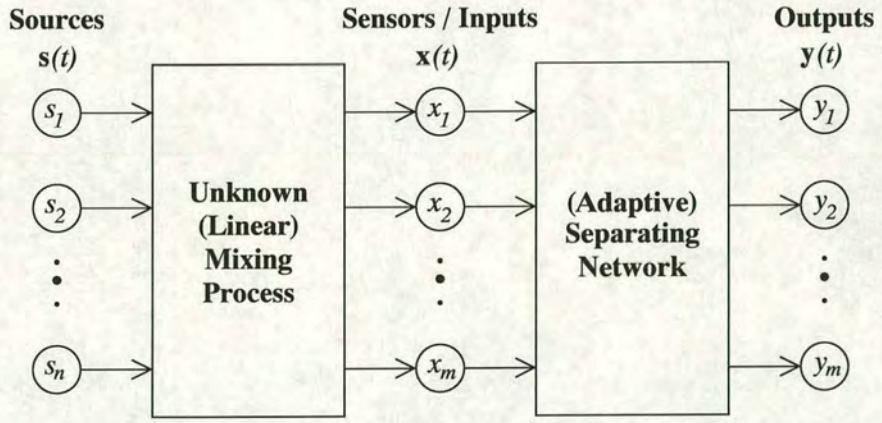


Figure 1.1: *Illustration of the blind signal separation problem*

simultaneous equations, but since the problem is under-constrained, the sources may not be recovered in the same order as the observed inputs, and may also be arbitrarily scaled. Whilst little can be done to map the order of the outputs to that of the inputs, this is generally not an important issue as it is the content of the outputs that is of interest. Furthermore, in most cases, the content is normally defined by the relative amplitude of the data described by the signals, and hence is scale independent. Consequently, these two uncertainties do not affect the validity of the outputs in any significant way.

1.1.2 Blind deconvolution

The blind deconvolution problem framework is similar to that of the separation framework, but instead of the instantaneous mixing matrix used to model the combination of the signals, a matrix of filters is used. Thus the mixed outputs now contain filtered, time-delayed images of the original signal as well. The solutions to blind deconvolution problems must therefore attempt to invert the filtering as well as reverse the mixing that occurred. This means that the approximate length of the filters to be inverted must be known in advance, as otherwise the inversion may either not be possible, if the estimated length is too short, or overly complex, if the estimated length is too long.

Care must be taken in the selection of the architecture and algorithm used to attempt such separation, as not all architectures and algorithms are capable of solving all potential types of mixing and filtering problems, or for all types of signal. In particular, some solutions will not work as well with signals containing any sort of temporal correlation, and others are only capable of dealing with minimum-phase filtering. As with the blind separation case, the outputs of a successful separation are subject to potential scaling and permutation.

1.1.3 Blind solutions

The methods used to solve the BSS problem take a variety of forms, many based around artificial neural networks (ANNs). ANNs are a class of adaptive computing units, each capable of modifying the function that they compute based on their current or past outputs, in order to 'improve' the results it produces in some way. This adaptability is normally achieved by defining the output function in terms of a set of weights which are iteratively updated, and which should eventually converge to some constant, final value assuming that one exists for the problem specified. In the BSS problem, the learning rule — which controls how the weights are updated — is designed in such a way that the weights should converge to values of an inverse of the mixing matrix that generated the input signals from the sources. In the blind deconvolution case, the weights should converge to the inverse of the mixing filters.

Convergence to the correct solution can be difficult if the sources have non-stationary characteristics, such as those exhibited by speech signals (at least, when being considered over periods greater than 20–25 ms). The definition of "stationarity" adopted in this thesis is that described by Papoulis [1] for *wide-sense stationarity* : a signal whose mean and variance

are constant. The description *non-stationary* may therefore be applied to signals that do not exhibit these properties. Non-stationary signals are an important category of signals due to the current proliferation of communication systems and applications for which they form the basis — and the need for speech processing techniques to make best use of the existing technologies and maximise the potential of future developments. The range of potential applications for many blind separation algorithms would therefore be increased if the algorithms could be used more effectively for separating non-stationary signals. In particular, methods for improving the separating performance of such systems for speech signals would provide useful extensions to the set of existing techniques. Other researchers, such as Torkkola [2–5], Van Gerven & Van Compernolle [6–8] and Nguyen Thi & Jutten [9], have considered other techniques to overcome this problem. Their approaches focused on pre-filtering or on controlling the update of the separating network based on the energy of the outputs, rather than the non-stationarity of the signals.

A variety of approaches exist for constructing suitable learning rules for the blind separation problem, and that proposed by Bell & Sejnowski [10], based on the information-maximisation (or infomax) framework was selected as the base approach for this thesis. The interesting information-theoretic approach adopted by the algorithm and its reported success with mixtures of stationary signals made it suitable for further investigation and development.

1.1.4 Information-maximisation

The information-maximisation framework approach to solving the blind signal separation uses a measure of the information content of the estimated sources at the networks outputs to determine how to update the system's weight set. It seeks to maximise the entropy of the output layer with respect to that of the input layer, whilst minimising the mutual entropy of the individual outputs. This work is based on earlier research by Linsker [11] into unsupervised learning in the visual cortex.

Whilst Bell & Sejnowski's framework [10] works well under certain assumptions, such as the independence of the source signals and that their characteristics are stationary, its performance suffers when the source signals of the input mixtures do not satisfy these restrictions. This poor performance on non-stationary signals, which, as previously noted includes speech signals, provides a starting point for this investigation.

1.2 Aims

The primary aim of this research is to determine the effectiveness of non-stationarity reduction by silence removal on the performance of blind separation and deconvolution techniques available for separating out individual signals from mixtures of speech signals. Given the plethora of voice-related applications in the communications field, a large and active area of research, there are many processing methods that can perform this separation to varying extents, albeit with some limitations on pre-requisite assumptions.

However, few of these methods can operate without some prior knowledge of the problem, such as the signal characteristics or the mixing process. Blind signal separation solutions, such as those described in Section 1.1.3, are important as these methods require less, if any, prior information and yet are normally able to provide superior separation performance. This capability enables them to be applied in situations where information about the signals or the mixing cannot be known in advance.

Even so, there are some choices to be made — certain systems are better suited to dealing with particular classes of signals. As described earlier, the information-maximisation approach works well with stationary source signals, but performs poorly on non-stationary data, a category into which speech signals fall due to their bursty nature. Of the active areas of blind separation and deconvolution research, those dealing with non-stationarity — whether in the creation of the mixtures, or in the signals characteristics themselves — have only relatively recently started to receive much attention. An assessment of techniques aimed at improving the separation performance of blind separation and deconvolution systems on non-stationary data will therefore provide a useful addition to that area of knowledge.

The infomax algorithm of Bell & Sejnowski was selected for use as the base approach in this investigation since it has been shown to perform well, albeit on sources with stationary characteristics, and does not directly require the estimation of higher order moments of the signals, as many other methods do. These estimates must usually be made over relatively long periods, and this cannot be done accurately for signals with rapidly varying non-stationary characteristics. In this study, 'rapidly varying' means with regard to the period over which the higher order moments must be estimated.

The thesis investigates the effect of reducing the degree of non-stationarity of the variance of speech signals as a possible means to improve the performance of blind separation and

deconvolution systems. The modifications investigated make use of techniques that reduce the short-term variance in the signals' amplitudes by identifying and removing periods of silence. These techniques are then incorporated into modified on-line separation systems. The main body of the work presented focuses on the information-maximisation approach for instantaneous mixtures, but experimentation is undertaken to evaluate and compare the effectiveness of the techniques when incorporated into other separation systems and deconvolution systems.

This project therefore investigated the hypothesis that non-stationarity reduction of speech signals by silence removal has a beneficial effect on the performance of blind signal separation systems.

A further introduction to each new topic is provided within the relevant chapters.

1.3 Thesis structure

The thesis structure is summarised here, for ease of reference :

- Chapter 2 introduces the key aspects of relevant background material, covering the underlying ideas of signal processing, adaptive filtering, blind signal processing and artificial neural networks (ANNs). It also discusses the suitability of using ANNs to implement solutions for blind signal processing problems.
- The blind separation and deconvolution problem are discussed in Chapter 3, as are applications of these techniques, and issues that must be considered. The remainder of the chapter focuses on existing neural and non-neural approaches to solving ICA, BSS and blind deconvolution problems.
- Chapter 4 describes the experimentation undertaken on instantaneous mixtures of speech signals. It examines the degree of non-stationarity of the signals, how this can be altered by various silence removal techniques, and the effect that this processing has on the separation performance of the information-maximisation algorithm, when applied to mixtures of these signals.
- Chapter 5 extends this work by developing the application of the silence removal methods into an on-line process. Alternative update strategies to augment the separation

performance are proposed and evaluated for various sizes of networks. Separation performance of these modifications is compared with that of other separation algorithms, and with those algorithms similarly modified.

- This on-line assessment is then applied to the separation of convolutively mixed signals in Chapter 6. This investigation considers both time domain and frequency domain assessment of the outputs, for filters with realistic characteristics.

- Chapter 7 presents a summary of the research undertaken, and the conclusions that can be drawn from them. It also identifies further areas of work that could be developed from the research presented in this thesis. More specific discussion and conclusions are presented in each of the chapters.

Copies of papers published from work undertaken in this thesis are reproduced in Appendix A, and statistical analysis of relevant data is provided on an accompanying CD — see Appendix B.

Chapter 2
Context

This chapter provides a general context for the research presented in the thesis by presenting a brief overview of the underlying principles upon which the research is built. An introduction to signals and signal processing is given, followed by a description of adaptive filters and artificial neural networks, two computational paradigms appropriate for the processing to be carried out. Finally some of the basic blind signal processing problems are set out.

More specific background on existing work in the field of the research can be found in Chapter 3.

2.1 Signals

A signal is a representation of information. A signal may be generated from a physical process by means of a sensor or a transducer, or generated artificially. Sensors usually output a signal that varies in amplitude in proportion to the intensity of the physical process they are measuring. Eyes and ears are examples of biological sensors, providing the brain with signals that it can then interpret and process. Microphones and antennas are examples of electronic sensors that output a time-varying voltage which can be processed elsewhere.

According to Balmer [12] a signal can be described mathematically as a function with a dependent variable and one or more independent variables, where one of the independent variables is normally time. The dependent variable is the value of the signal at the specified time.

If a signal is a complete translation of every variation of the process at every instant of time, it is said to be *continuous*. If it is formed as a series of regularly-spaced brief observations of the process, then the signal is described as *discrete*. The production of a discrete signal from a continuous process or signal is performed using a technique known as sampling. The (regular) time interval between successive observations determines the sample rate, which should be made fast enough to ensure that all of the desired changes within the signal will be recorded. If

the sampled signal can only take on specific values within a defined range, it may be described as *digital*, otherwise, if the permissible values are continuous and unconstrained, it is said to be *analogue*. The digital representations of signals are often encoded to simplify the processing of the signals.

2.2 Signal processing

Whenever a signal passes through any system that modifies it, it can be said to have been processed by that system. The system may be a natural phenomenon such as a physical medium through which the signal is propagating, or a designed device such as a filter or a computer chip. The processing carried out may be a simple, fixed operation such as a re-scaling of the signal's amplitude, or a complex sequence of operations that may vary with time, or with certain characteristics of the signal or system.

Common applications of signal processing include those concerned with electronic representations of audio or visual data. These are frequently encountered in home entertainment systems, communication activities and medical analyses. Typical signal processing activities include the transformation and conversion of the signal data from one representation or format to another, amplification of the signal, filtering to remove noise or specific parts of the signal, and analysis of the new signal characteristics.

Since the signals can be represented electronically and thus processed using a computer, the range of processing that can be performed is far greater than was ever possible using analogue devices. The continual increase of computing power facilitates more and more complex processing, allowing the application of advanced techniques to solve traditionally difficult problems. This opens up opportunities for research in areas that were not practical, or possibly not even recognised, before. These advances include techniques and applications based around signals from arrays of sensors, which often require a geometric expansion of processing power to deal with the multitude of signals. Sensor arrays offer the potential for enhanced processing, as they provide multiple images of the signals received at any instant in time. These multiple images can greatly assist the processing of signals by providing additional information and, with knowledge of the array layout or structure, can be used to even greater effect.

Applications of sensor arrays included exploratory or investigative tasks such as seismic surveying, radar- and sonar-based techniques, beam-forming, medical imaging and

communications.

2.2.1 Adaptive signal processing

In many cases, the exact processing to be performed on a signal can be fully specified in advance. However, for systems that must deal with real-world signals or phenomena, this is not always possible as the exact conditions or characteristics of the signals or phenomena may not be known in advance or be accurately predictable. In these cases, some degree of flexibility is required in carrying out the processing. This may be as simple as accommodating a time-varying range of values rather than assuming a predetermined fixed level, or as complex as altering the processing carried out according to not only the current value of the input signals, but also their previous and possibly even predicted future values.

Processing that is capable of dealing with such variation is described as *adaptive*. Adaptive processing is usually more complicated to define than non-adaptive processing as it must take account of the variability of the signals or system. The adaptation may be open-loop or closed-loop. Open-loop systems endeavour to maintain a particular relationship between the inputs and the system, whereas closed-loop systems offer a self-optimisation capability usually involving a feedback mechanism based on the system's outputs. In the feedback loop, the results from the current processing are used to update the system in a way that should improve the system's performance. Care must be taken with closed-loop systems, as errors in their outputs may lead to incorrect adaptation, resulting in divergence from the desired solution.

Hence, as well as defining the processing that must be carried out in adaptive systems, there is usually a causal, temporal relationship that defines how this processing is to be updated. A simple model of this basic arrangement is presented as Figure 2.1. Over successive iterations

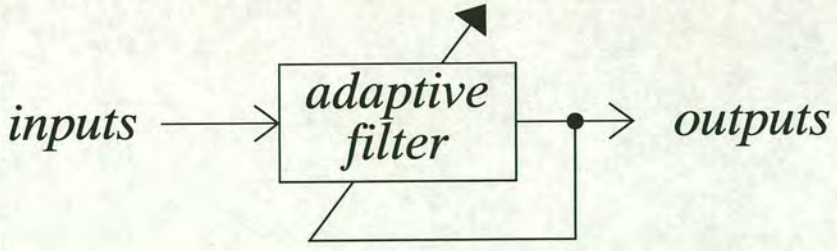


Figure 2.1: *Adaptive processing*

of this loop, the performance of the system should converge to an optimal value, subject to the

characteristics of the inputs and the rest of the system remaining the same. This means that adaptive systems are capable of tracking non-stationary characteristics, as long as they change only slowly with respect to the update period.

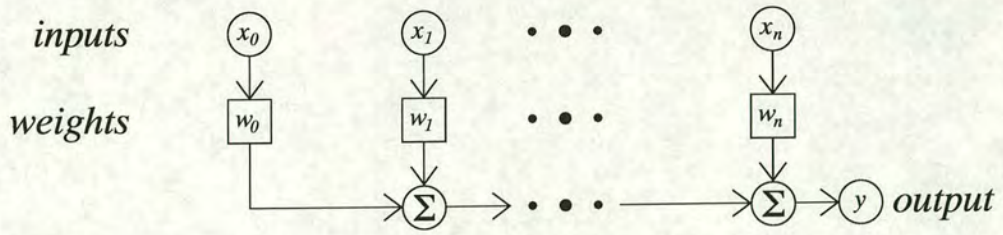
To satisfy the requirement of being able to update the processing, the architecture of adaptive systems must incorporate an appropriate mechanism. Usually, this separates the values used in the definition of the processing from the processing itself, leaving the latter fixed. There are many possible architectures that support this, two of the most common being described below.

2.2.2 Adaptive filters

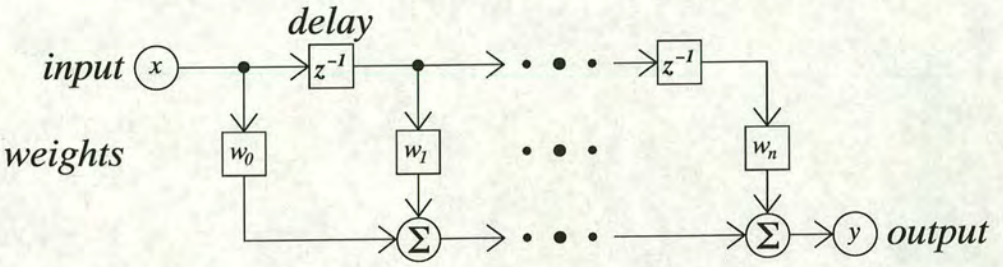
A non-recursive digital filter whose tap weights can be modified is an example of a simple form of an adaptive filter. The values of the tap weights at the time of the processing define the exact proportion of the data values that are combined in the way defined by the structure of the filter. The input data to the filter can either be from separate input signals, or from successive samples of a single signal. In the latter case, each pair of samples is assumed to be separated by a delay element in the filter structure, which is then known as a transversal filter or a finite impulse response (FIR) filter. These two cases are shown in Figures 2.2(a) and 2.2(b). More complex structures, such as lattice configurations, are also widely used but are not dealt with in this thesis.

In order to drive the adaptation, an algorithm or *learning rule* is required. It is this algorithm that determines the adjustment to be applied to each of the weights which make up the filter. Error-correction rules are one of the most basic forms of learning rule, based on the method of least squares. Haykin [13] describes an approach to the development of adaptive algorithms built around this technique. The weights of a filter are updated in relation to the distance between the desired output, or some property of this, and that actually calculated by the filter. This measure, the error, must usually be minimised to achieve the desired solution. These rules are sometimes also referred to as *delta rules*, as the changes to an individual weight w_i are denoted Δw_i. The stepwise adjustments are calculated to move the output approximation closer to the desired output, the size of the step being controlled by a learning rate parameter, normally denoted η. The development of a particular weight i can thus be encapsulated by Equation 2.1 :

$$w_i\,(t+1) = w_i\,(t) + \eta \Delta w_i\,(t) \tag{2.1}$$



(a) Multiple input adaptive filter



(b) Single input transversal filter

Figure 2.2: *Simple adaptive filter architectures*

where $w_i(t+1)$ and $w_i(t)$ are the new and old values of the weight, respectively.

Haykin [13] also describes the development of another set of algorithms, this time based on stochastic gradient approaches. Here, the current values of the weights are used to estimate a position, on the surface defined by a cost function, in the weight space of the network. The update of each of the weights is normally carried out so as to move the position of the network's current solution on this surface either uphill or downhill, according to the learning rule. This can be achieved by differentiating the cost function at a point defined by the weights, which determines the gradient of the surface local to that point. The size of the step may be limited by using a learning rate parameter as before.

With some systems, the network may become trapped in local maxima or minima of the function's surface, from which it may yield only poor solutions. A variety of techniques such as simulated annealing [14] can be used to escape from such extrema. Such 'trapping' may not be such a problem in dynamically changing environments, where the surface can vary with time, as maxima and minima may form and disappear during the evolution of the system — but this will depend on the type of non-stationarity in the system and the form of the cost function

defining the surface.

Adaptive filters are employed in applications such as adaptive equalisation, adaptive noise cancellation, and adaptive beamforming (these being distinct from their non-adaptive counterparts).

2.2.3 Artificial neural networks

Artificial Neural Networks (ANNs) are another class of computational processing architectures that are capable of adaptively adjusting their computation. They are employed in systems to carry out such tasks as statistical classification, pattern recognition, function approximation, image analysis, prediction, process control and signal processing.

The "neural" aspect of the name relates the manner in which the computation is updated to the learning capabilities of biological neurons. ANNs are connected arrays of relatively simple computational processing units referred to as *neurons*, or *nodes*. Each node in such a network is generally capable of performing only relatively simple calculations, normally a sum or product function of its inputs, which may then be processed by a non-linear function. There are a variety of architectures, but most conform to the generic structure shown in Figure 2.3, which is similar to that of Figure 2.2(a) except that it is followed by a non-linear transformation. Equation 2.2

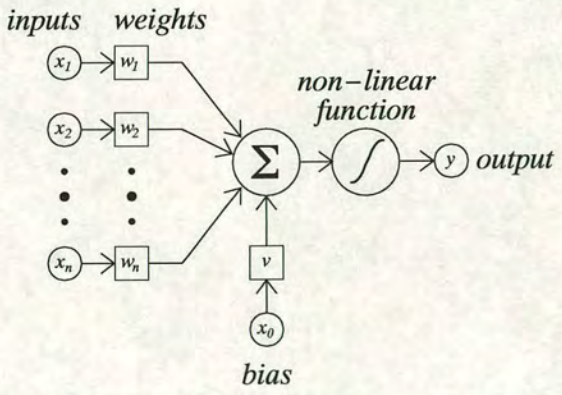


Figure 2.3: *A typical node structure*

gives the formula for calculating the output, y of a n-input, single output node conforming to this structure, with input vector, $\mathbf{x}$, synaptic weight vector $\mathbf{w}$, bias term v and non-linear

function $f()$.

$$y = f\left(\sum_{i=1}^{n} w_i x_i + v x_0\right) \tag{2.2}$$

The real power of such systems comes from the synergy of the composition and interconnection of these units into layers and networks, as their computational power is cascaded and thus multiplied. Continuing the biological analogy, the connections between neurons are described as synaptic and have some weighting (scaling factor) ascribed to them, the value of which indicates the importance of that connection in the calculation of the (sub-)network output. Connections may be feedforward, lateral or feedback, as illustrated in the generic network model of Figure 2.4. Adaptation occurs only within the stored set of weights, which are used in

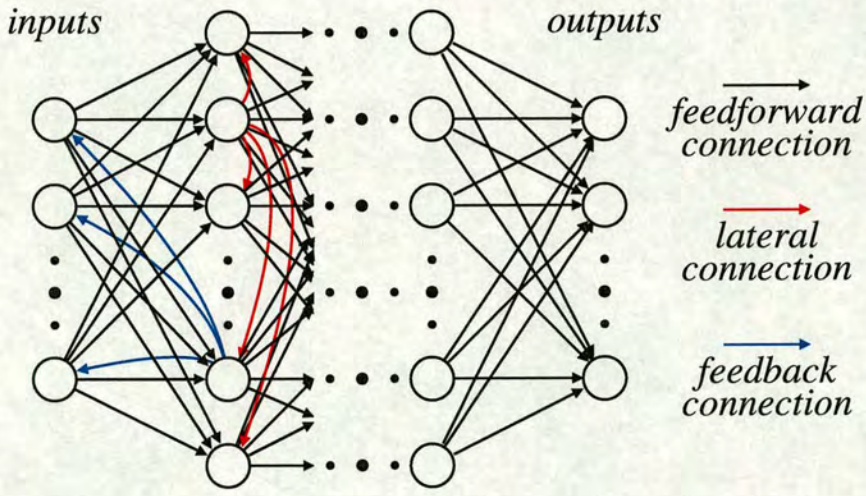


Figure 2.4: *A typical artificial neural network architecture*

determining the system's state. This is then used to calculate the new outputs from the current inputs. These new outputs are, in turn, used in calculating a set of updates for the weights, and hence there exists a feedback mechanism by which the system adapts or "learns".

2.2.3.1 Learning strategies

The learning strategies control how the weights of the network are updated in response to the outputs and the current network state. There are even more learning strategies than there are architectures for ANNs, but they can be divided into a number of distinct categories, some of which correspond to those used by adaptive filters. The first division is based on whether

or not the updates can be independently calculated by the network itself, or whether external assistance is required to do this. Those networks that require external assistance are called *supervised* networks, and correspondingly, those that do not are called *unsupervised* networks.

Supervised networks The external assistance normally takes the form of a 'teacher' or 'monitor' that analyses the network's outputs and determines how well the network is performing. The assessment is then used in controlling the update of the network weights — normally based on error-correction learning (see Section 2.2.2). Once a desired performance level has been reached, the teacher can be switched off, and the system left to run on-line. This means that the level of performance of the network is now fixed, assuming that the characteristics of the inputs do not change with respect to the way in which the outputs are determined from them.

Unsupervised networks Unsupervised networks do not require a separate teacher or monitor to assess their outputs, or to tell them how to adapt. Instead, the update-generating rules are formulated in such a way that the updates are defined with respect to some property of the desired solution, or a task-independent measure. Their autonomy makes unsupervised networks the ideal choice for situations where the output is not — or possibly *cannot* — be known in advance. This allows processing of unknown or undefined signals, where the aim is to modify the signals in some way, rather than to achieve a specific output.

The learning rules for these networks are normally posed either in terms of some constraint, such a limit on the range of values of the weights, or as a cost function. The cost function normally computes a value based on the state of the network, which must be maximised or minimised, either globally or locally, for each of its nodes.

Since the calculation of these updates can be computationally intensive, the computational load of repeatedly updating the network can sometimes be reduced by using *batch processing* techniques. This involves performing a lesser computation at each iteration (usually just the addition of the current inputs or weight set to a running total of that value) and then carrying out the complicated transformation of the update rule on the sum of b iterations of values (where b is the batch size). In this way, the computational load is reduced since the intensive update calculation is only carried out $\frac{1}{b}$ times as frequently. An output from the network is still calculated at each iteration, but now b of them will use the same values of the weight set.

2.3 Blind signal processing

So-called "blind" signal processing is a branch of adaptive signal processing that is further complicated by the fact that there is no *a priori* knowledge of the signals involved. This means that the processing must be defined in fairly generic terms since no specialised techniques, tailored to particular forms or types of signal, can be applied. It is also what gives rise to the description of such processing as "blind" — the adaptation must be carried out blindly, *i.e.* without knowledge of the desired outputs, even though the desired result is known in advance — *e.g.* the separation of the signals.

In developing solutions to blind signal processing problems, it is common to either relax the "blind" constraint and allow some assumptions to be made, or to partition the problem into separate, smaller problems, each constrained to a restricted range of conditions. These restrictions provide additional information that can be used, in the same way as the assumptions in the alternative approach, to form a basis on which to build (or optimise) a solution. The assumptions or restrictions need not be related to the signals being processed, although they often are. Instead, they may pertain to the system generating the signals, or to certain conditions of the problem specification.

Many of the classic blind signal processing problems involve multiple observations of one or more signals — in fact, the multiplicity is precisely what makes some of these problems blind. Consequently, solutions to them often involve the use of sensor arrays or sampled data systems. Some of these classic blind signal processing problems are described briefly below, along with comments on solutions or methods of tackling them. Their descriptions identify both the problem and the name given to its category of solutions :

Transmitter location — direction of approach (DOA) The problem here is to determine the direction (relative to the receiver array) from which the signals of interest are being broadcast, as described by Farina [15]. The approach taken by Burel and Rondel [16] makes use of the relative spacing between sensors in the receiver array and the rate of propagation of the incoming signals to determine the phase difference between the sensors, and hence deduce the angle that must give rise to this offset.

Channel distortion — blind equalisation Blind equalisation solutions attempt to determine a set of filter coefficients required to restore a received signal that has been distorted by channel propagation effects during its transmission. Solutions to this problem include

the use of sequences of symbols, and estimations of the entropy of the source.

Signal convolution — blind deconvolution The intention of blind deconvolution is to eliminate time-correlated echoes of a signal, leaving just the original. This involves identifying the original signal and constructing a filter to combine appropriate echoes in a constructive manner. This problem is not limited to the audio domain, although it is is perhaps best known here for the so-called "barrel-effect" often encountered with speaker-phones, or for the stereotypical announcements at train stations.

Signal mixing — blind separation Blind separation aims to separate m linear mixtures of n statistically independent signals that have been linearly mixed, back into their constituent sources. There are many approaches to solving this problem, based around either higher order statistics or information theory. Blind separation is one of the problems addressed in this thesis, and hence more details on both the problem and some of the solutions to it are given in Chapter 3.

Combinations of such problems can also occur — for example, the mixing of a set of convolved signals. Consequently, there is also a need to solve these combined problems. In such combinations, the difficulties of the individual problems often compound one another. Therefore it is not usually the case that the solutions to the problems can be simply combined. The specific problem of the mixing of a set of convolved signals is also investigated by this thesis — details are given in Chapter 3.

As well as dealing with the unknown elements of the blind signal processing problems, the generic techniques that must be employed must also take account of the fact that the signals being considered are often time varying. As a result, the solutions to blind problems are usually based around adaptive algorithms. In particular, Artificial Neural Network (ANN) techniques are frequently applied to these sorts of problems, for the reasons previously described (see Section 2.2.3). In these situations, the unsupervised, constraint-driven configurations are more appropriate, with the inputs being the signals to be processed and the learning rules being defined in such a way as to result in the network's processing producing the desired effect on the outputs. The development and assessment of such techniques form the basis of the research in this thesis.

2.4 Summary

Having introduced adaptive filters and artificial neural networks, in the context of adaptive signal processing, the use of the latter to carry out unsupervised adaptive processing has been proposed as a suitable framework under which to tackle blind signal processing problems. The learning rules for these networks can be developed such that they incorporate properties of either the problem or the solution. They thus update the network's weights in a manner leading to their convergence to values that result in a solution to the problem. This framework forms the basis of the approaches taken in the research presented in this thesis.

Chapter 3
Background

Having set the context of this thesis and provided required background information in Chapter 2, this chapter examines the blind separation and deconvolution problems in more detail, their applications in different fields and implementation issues. Considerations above and beyond the strictly blind constraints are also addressed for both general and specific variations of the problems. Summaries are then presented of relevant approaches that have been taken to solving the blind signal processing problems, covering both neural and non-neural solutions, and the aims of the thesis are re-stated on the basis of this background information.

3.1 Blind separation and deconvolution

Blind separation and deconvolution are two of the so-called "blind" problems, the characteristics of which were described earlier in Section 2.3. Both are important signal processing problems which are being actively explored, and various solutions have been developed for application in a number of diverse fields (see Section 3.1.3). This thesis considers the blind separation and the combined blind separation and deconvolution problems, which are described below, along with examples of how some of the solutions have been used in different disciplines. The blind deconvolution problem on its own is not directly discussed.

3.1.1 Problem specification

Although distinct problems in their own right, the class of problems requiring a combination of blind separation and blind deconvolution presents a much more realistic model of the observed effects of the physical process of mixing in a real-world environment. To provide a more detailed insight into issues of these combined problems, the blind separation problem is initially considered on its own.

3.1.1.1 Blind separation

There are a number of subtly different problems that are often commonly referred to as "blind separation". These include Blind Signal Separation, Blind Source Separation and Independent Component Analysis (ICA). While the result of solutions to these problems is the separation of the mixed inputs signals, the perspective of the solutions can be different. Some, such as ICA, may place a greater emphasis on the determination of the separation system, while others focus on the signals themselves. Blind source separation (BSS), for instance, specifically considers the case where the m signals to be separated are defined as being formed by linear combination of n (unknown) independent source signals, without any *a priori* knowledge of these sources, or their mixing. The observable mixtures are the outputs from a sensor array (often an array of microphones), and the source signals are realisations of some physical process, such as speech. It is this specific problem of BSS that is considered in this thesis.

Blind separation can be carried out on both instantaneous and convolutive mixtures. The general situation for instantaneous mixtures is considered first, and can be illustrated by the diagram given in Figure 1.1, reproduced here as Figure 3.1 for ease of reference. It is

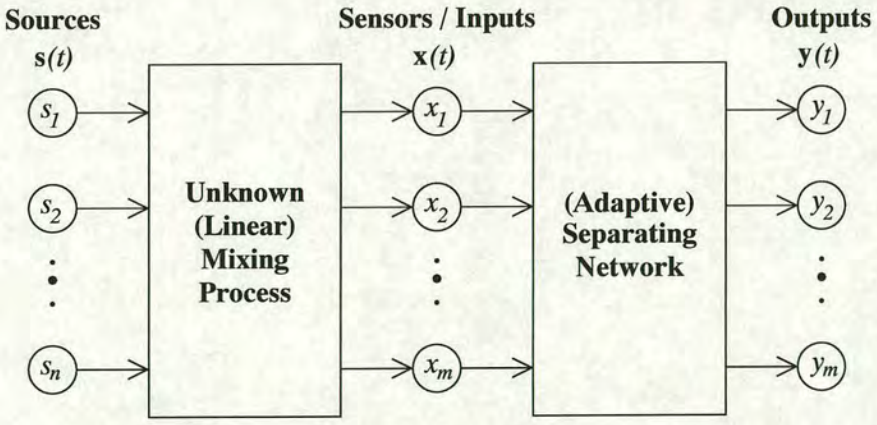


Figure 3.1: *An illustration of the blind signal separation problem*

exemplified by the classic *cocktail party problem* [17], which is that faced by someone walking into a room where there are many conversations going on simultaneously. A human listener can, with relative ease, filter out the background noise and other conversations, and focus in on one conversation of interest — however, to create an artificial system that can perform this same task without some prior knowledge of the speakers' characteristics, is considerably more difficult.

To simplify further explanation of the approaches taken, the notation that will be adopted for the remainder of this thesis is presented in Table 3.1. This differs slightly from that found in some of the references, but is an attempt to unify the range of different formalisations used. A mathematical view of the problem is adopted as this allows concise expression of the terms involved. Subscripts are used to denote individual elements of vectors, such as x_i or y_j, and the

s	The vector of independent source signals
A	The (non-singular) mixing matrix
x	The vector of linearly mixed input signals
W	The (approximated) separating matrix
y	The vector of estimated sources as outputs

Table 3.1: *Mathematical notations for BSS / ICA, used in this thesis*

standard row-column ordering is applied to matrices, so that $\mathbf{w}_{ij}$ represents the element at the j^{th} column of the i^{th} row of matrix $\mathbf{W}$, all indexing starting at 1. Since the signals of interest are time-varying, a snapshot of a particular signal (*i.e.* one sample) is represented by the use of a time index, such as $x_i(t)$. In the absence of this time index, the whole of the signal is considered, unless otherwise stated.

The mixing process of the problem specification can now be defined by the relationship :

$$\mathbf{x} = \mathbf{A}\mathbf{s} \tag{3.1}$$

where $\mathbf{s} = [s_1, s_2, ..., s_n]^T$ and $\mathbf{x} = [x_1, x_2, ..., x_m]^T$. The aim of BSS is to reverse this mixing, given only the set of observed mixed inputs $\mathbf{x}$, and determine a separating (or 'unmixing') matrix $\mathbf{W}$ such that :

$$\mathbf{y} = \mathbf{W}\mathbf{x} \tag{3.2}$$

where $\mathbf{y} = [y_1, y_2, ..., y_n]^T$ is a (possibly scaled and permuted) estimate of the source signals $\mathbf{s}$. Due to the under-determined nature of the problem, which gives rise to the arbitrary scaling and permutation factors, this is normally qualified such that :

$$\mathbf{C} = \mathbf{W}\mathbf{A} \rightarrow \mathbf{D}\mathbf{P} \tag{3.3}$$

Here $\mathbf{D}$ is a non-singular diagonal matrix that can account for the scaling of each of the outputs,

and $\mathbf{P}$ is a permutation matrix (only one non-zero element per row and column). The matrix $\mathbf{C}$ characterises the whole system, eliminating the emphasis of particular values of the mixing matrix and scaling factors. Without scaling and permutation, it would converge to the identity matrix $\mathbf{I}$.

The separating process endeavours to update the values of the separating matrix $\mathbf{W}$ so that they converge to a solution where the outputs $\mathbf{y}$ are all statistically independent (or in a practical application, as independent as is possible [18]). Since this must be done without knowing what the 'correct' value of the outputs are (a so-called "unsupervised" system — see Section 2.2.3.1), the updates must be determined by the use of a constraint or a learning rule based on one or more properties or characteristics of the desired solution. When a value can be assigned to the degree to which the current system conforms to the desired property or characteristic, the constraint or rule is referred to as an *objective function*. Maximisation or minimisation of the objective function should correspond to the fulfillment of the output criteria. As described previously in Sections 2.2.2 and 2.2.3.1, techniques exist for iteratively achieving such maximisation or minimisation in an unsupervised manner.

The blind separation of convolutive mixtures shares many common features with the combined blind separation and deconvolution problem, and hence is discussed in the following section.

3.1.1.2 Blind separation and deconvolution

While the blind separation problem provides a convenient way of expressing or visualising the concepts involved in the separation of mixed signals, it does not present a particularly realistic model of the mixing that takes place in a real world environment. Here, there are additional factors to consider, such as delays between the sources and the sensors, and the effects of reverberation (echos) on the received signals. These delays and echos can be modelled by filtering the source data.

Some of the blind separation techniques can be extended to accommodate the multiple samples implicit in the filtering operations, and used to separate the individual signals. However, the outputs are now subject to arbitrary filtering, in addition to the arbitrary scaling and permutation, leaving them distorted compared to the original signals. Extending the blind separation problem to take account of the filtering effects as well requires solving a combined blind separation and deconvolution problem.

The solution to the basic blind deconvolution problem considers trying to filter a signal to remove the super-imposed time-delayed images of the signal, introduced by the system response of some influencing factor — be this the transmission medium (channel) through which the signal is propagating, some artificial processing, or whatever. When extended to be a multi-channel situation, where cross-talk between the channels is possible, the model is that of the combined blind separation and deconvolution problem. Here, the two problems are intricately linked and the observed signals contain not only the time-delayed images of themselves, but are now also contaminated with some proportion of the other signals and their images as well. The aim of the various techniques is still the same, however — to recover the original signals (and in some cases, also the set of inverse filters).

The mathematical framework proposed in the previous section can be extended to accommodate this added complexity. The two-dimensional mixing and separating matrices ($\mathbf{A}$ and $\mathbf{W}$ respectively) are extended into three dimensional arrays, making them both arrays of filters. The array of mixing filters is usually denoted $\mathbf{H}$. The original two dimensions of the array are the same as before, number of sensors (n) and number of sources (m), and the new third dimension is the time delay, which shall here be denoted by l. Again, its value would be unknown in a truly blind situation. The time delay is indexed by the tap number of the filter, each tap corresponding to successive samples of the digitised signal. Individual array elements will now be identified as w_{ijk}.

Where instantaneous mixing of the signals was mathematically represented by matrix multiplication, the temporal aspects of the mixing and filtering of the signals must now be represented by the *convolution* of the source signals with the array of (mixing) filters ($\mathbf{H}$) as :

$$\mathbf{x} = \mathbf{H} * \mathbf{s} \tag{3.4}$$

where $*$ is the convolution operator for time domain signals, and the other terms are as in Equation 3.1. Similarly, the solution sought is now :

$$\mathbf{y} = \mathbf{W} * \mathbf{x} \tag{3.5}$$

The scaling and permutation indeterminacies still exist, such that the entries of

$$\mathbf{C} = \mathbf{W} * \mathbf{H} \rightarrow \Lambda\mathbf{DP} \tag{3.6}$$

contain one scaled, permuted, and now potentially independently-delayed impulse response in each column or row of the first two dimensions of the arrays. Λ is an array of delays that offset the impulse responses in each of the non-zero filters, allowing non-causal filters to be rendered as causal. $\mathbf{D}$ and $\mathbf{P}$ are as before, but now applied to each filter in the array as an entity. In the blind separation of convolved sources, this delay would be replaced by the arbitrary filtering operation.

The separation process then proceeds as before, the network driving the iterative update of the weight set $\mathbf{W}$ such that the outputs $\mathbf{y}$ are as independent as is possible, and in the case of the combined separation and deconvolution solutions, are also deconvolved.

3.1.2 Other signal separation work

Whilst other ways of extracting signals from mixtures exist, most work on only single sensor outputs and are often not classified as signal separation since they simply suppress the undesired part of the mixture. The desired parts of the mixture have not been separated, as such, and in many cases separation has not even been attempted. The discarded parts may thus still contain potentially useful signals. Furthermore, information about each of the desired signals' characteristics is often required in advance, to allow the extraction process to be tailored to the particular signal in question. As a result, these approaches cannot generally be used directly to solve the blind separation problem, and consequently, are not further considered.

3.1.3 Applications

There are many potential applications for blind signal / source separation or independent component analysis based systems. While effective, these are limited by the current knowledge of how to handle additional complications such as the presence of noise in the signal mixtures, the linearity of the mixing process and an unknown number of the source signals.

Principal problem areas to which these techniques have been applied include :

Antenna array processing Antenna array processing is one of the most obvious systems to which BSS or ICA can be applied, since the multiple sensor data is inherent in the very nature of the work. Data from radar and sonar arrays have been considered, and clear improvements in data separation have been achieved [19–21].

Communication There are a number of areas in the wireless communications field that benefit from BSS / ICA techniques [22] — techniques have been used to perform noise reduction and signal enhancement on Orthogonal Frequency Division Multiplexing (OFDM) [23] and Code Division Multiple Access (CDMA) systems [24], to name but two.

Speech separation This is frequently used as a demonstrator application, since the data can be assessed relatively easily and compared to the original signals and to the mixtures. However, speech separation is also an important task in a number of real-world situations, separating speech either from other voices, such as in a teleconferencing scenario, or from background noise. Whilst this latter target could perhaps be achieved by other filtering techniques, some of the blind signal separation approaches offer additional advantages that the others cannot. To this end, speaker identification, speech recognition and some hearing-aid related tasks could all benefit from these BSS techniques. Speech separation is a key part of [10] and work by [17,25] is also based on this. It is also the focus of this thesis.

Biomedical data analysis Many biomedical applications can make use of multiple sensors in cleaning-up crosstalk or interference from other corrupting sources in data. Particular uses of BSS algorithms that have been the subject of interest are magneto- and electro-encephalographic data analysis [26–28], and foetal heartbeat monitoring [29].

Condition monitoring Some pieces of machinery, particularly those involving rotating parts, exhibit characteristic signatures, in terms of signals of specific pitches and harmonics. ICA can be used to identify potential problems with the machinery by enabling the detection of subtle changes in the data recorded by monitoring equipment. Isolating the changes may allow localisation of the problem, reducing the time taken to trace the fault, and may even indicate the type of fault, if it produces a distinctive change in the signature [30,31].

Financial data analysis ICA has been used by Back [32] to try to determine the underlying trends in stock-market data, based on multiple observations of share prices across a range of trading categories. Such an analysis cannot hope to separate *all* of the independent factors that control the trends in the market, but instead focuses on identifying those components that exert the greatest influence.

Data compression It has previously been shown that ICA can be applied to images to perform feature extraction, yielding a set of components that constitute the image. Identification

of these components may allow for a more efficient encoding of the image than even that produced by Huffman coding. The principles may also be extended to other types of data, yielding better data compression rates than standard techniques [33,34].

Data analysis Yet another field to which ICA has recently been applied is that of unsupervised exploratory statistical analysis of data, or data mining [35]. In such analyses, an extended version of the infomax algorithm based on maximum likelihood estimation (MLE) is used to hierarchically partition clusters of data from multivariate distributions.

This is by no means an exhaustive list — many other application areas exist. However, the above list illustrates the wide range of the *types* of problem which can benefit from BSS or ICA systems.

3.1.4 Implementations

Whilst it is true that for many off-line post-processing cases, signal separation can be carried out by software running on a computer, this is not always convenient for on-line processing. For example, if the separation system was to be a part of a high-speed data pipeline in an airborne radar system on a fighter aircraft, then additional dedicated hardware optimised to cope with the rate and type of incoming data may be required. Factors that may lead to the rejection of computer-based solutions include size and portability, where a small, hand-held or possibly even single-chip device would be more practical — such as for use in hearing aids, or remote sensing systems. These units may also be more cost-effective, and certainly in terms of performance versus cost per unit, far preferable.

Until recently, there had been relatively few hardware implementations of blind signal separators or independent component analysers, especially given the amount of interest in the field. Most of the research carried out to date has therefore been based on the results of simulations, or on other software modelling of the problem. One of the major factors that originally deterred commercial interest in the field is likely to have been the computational complexity of some of the algorithms.

High complexity is not required for a basic, functioning system — see Section 3.2.5. Some of the newer solutions have been designed with digital signal processor (DSP) based systems in mind [36], but analogue VLSI chips, based on the simpler rules, have also been fabricated [37,38]. Presently, many of the modern DSPs and embeddable processors are more than capable

of dealing with the levels of processing required for blind separation systems — even those using some of the more complicated algorithms.

A consideration to be borne in mind when deciding on which system to implement, and in what technology, is that of the difficulty of designing structures for performing mathematical operations in analogue VLSI (aVLSI). The use of aVLSI would probably lead to the most efficient power / area usage, although it may be outweighed by the advantages of off-the-shelf DSPs for ease of design. Reconfigurability is another issue that must be addressed, since generic algorithms to deal with undefined numbers of sources are are also being investigated [39]. Consequently, network sizes may have to be changed either between runs of the process, or during its operation. Whilst this would be possible with some complex design of DSP systems, Field Programmable Gate Arrays (FPGAs) or embedded processors would offer much more flexibility in this respect. In terms of the algorithms selected for implementation, those which make use of local learning rules, and that avoid complex mathematical operations clearly offer advantages in terms of lower computational loading.

3.2 Issues in blind source separation

As previously described, a number of variations of BSS solutions exist to deal with different aspects of the problem. Most of these differences are concerned with either the mapping between the source data space and the mixed data or the characteristics of the data, and give rise to a large number of algorithms and techniques that have been developed to specifically exploit the peculiarities implicit in those differences. The different solutions are often based around alternative views of the BSS problem, some of which may give more insight into particular aspects of the situation and permit certain simplifying assumptions to be made. Many excellent reviews of some of the different perspectives and approaches exist, including those by Karhunen [40], Lacoume [41], Amari [42], Oja *et al.* [43], Girolami [44] and Torkkola [45].

These differences allow the broad range of solutions to be categorised in a variety of ways. The categorisations that will be considered here are the stationarity of the signals and of the mixing, the temporality and linearity of the mixing, and the generality and complexity of the algorithm. These divisions are based on those identified by Lacoume in [41].

3.2.1 Stationarity

The term *stationarity* is used to described whether or not certain characteristics of the factor under scrutiny change over time. If the characteristics do not change, the factor is described as stationary, otherwise it is said to be non-stationary.

3.2.1.1 Signal stationarity

When assessing signals, the stationarity of the moments of the signal is used to classify the signal as a whole. Frequently, only the first and second moments (the *mean* and the *variance* of the signal) are considered in this assessment.

Stationarity assessments need not be carried out over the whole of a signal at once — rather, they may be performed over a short section of the signal. Some signals' characteristics may be classified differently depending on the duration of this assessment window. This is true of speech signals. When considered in periods of 10–20 ms, speech is usually classified as stationary. When the periods in question are above this threshold, the characteristics tend to be non-stationary, due to the "bursty" nature of the signals.

3.2.1.2 Mixing process dynamism

As well as the characteristics of the signal data being potentially non-stationary, the mixing process that transforms the source signals to the input signals may also be dynamic, rather than static. Dynamic mixing environments are sometimes described as non-stationary, although this is not strictly accurate. In the dynamic case, the mixing that occurs will vary over time. This could come about as a result of the relative movement of one or more of the sources or sensors, or from something passing between the sources and the sensors, thus changing the mixing environment.

3.2.1.3 Stationarity considerations

As noted previously in Section 3.2, the set of algorithms capable of dealing with non-stationary signals or dynamic mixing environments can be assigned to a separate class. These methods must be capable of adaptively tracking changes in the characteristics of the mixed data, or in the mixing environment. Separation is considerably harder than for the stationary or static cases,

as the convergence to a stable solution must continually take account of the non-stationarity or dynamism.

This means that whilst these techniques may make some use of batch processing optimisations, these must be constrained by parameters such as the rate of change of the time-varying values. The learning rate, η, used to limit the maximum change that can be applied to update the system at each iteration must be set at an appropriate level that accommodates the changes due to the variation without allowing the approximated solution to diverge from the intended target. Anemüller and Gramss [46] adaptively update this parameter and an estimate of the signal energy, in a manner not dissimilar to that used later in this thesis. However, their focus was on a dynamic mixing environment, rather that the static one considered in this study.

Work on dynamic (non-stationary) environments has also been addressed by Naudet [47], Parga and Nadal [48] and Everson and Roberts [49]. Other research on temporal aspects of blind separation, such as the context-sensitive ICA methods as proposed in [50] have been investigated. However, since these situations are not to be considered in this thesis, the work is not discussed further here.

Other work on non-stationary signals that is relevant to this thesis is listed below. For more detail on these approaches, please refer to Section 3.4 :

- *Time dependent learning rules* Matsuoka *et al.* [51], [52] incorporate the signal non-stationarity into the objective function used in the network's adaptation.
- *Pre-filtering* Barros and Ohnishi [53] filter the non-stationary input signals before attempting to calculate each update.
- *Nonholonomic orthogonal learning constraints* Amari, Chen and Cichocki [54] constrain the updates generated from the non-stationary data to permit only those that will not lead to unstable solutions.

Of these approaches, the first two are considered in the framework of this thesis. The latter, although effective, is omitted due to the complexity of the updates (see Section 3.2.5).

3.2.2 Temporality

The term "temporality" is used here to indicate the distinction between solutions that depend on values at only a single instant in time, and those that depend on values at several different points in time. These temporal concerns are usually related to the mixing of the source data to produce the input signals. There are two possibilities :

- *instantaneous* mixing, in which the outputs are solely a function of the inputs at that moment
- *convolutive* mixing (filtering), in which additional combinations of past inputs must also be considered in determining the outputs

It is these two different classes that give rise to the distinction between the blind separation and blind deconvolution approaches previously described. It is not usually possible simply to extend an instantaneous solution to tackle a convolutive mixing problem merely by replicating it for all of the different points in time. Consequently, the solutions to the convolutive problems are usually more complicated their instantaneous counterparts.

Separation and / or deconvolution is only possible in cases when the mixing is invertible. In the case of instantaneous mixing, this implies that an inverse of the mixing matrix exists. For convolutive mixing, it means that the inverse of the mixing filters must be realisable. Thus there is the additional constraint of the causality of the separating filters to be considered — any filter that attempts to counter a delay in the formation of an input signal will, by definition, be non-causal. Techniques exist, however, to assist in cases when the inverses of the mixing filters are non-causal. These involve shifting the filter taps to make the filter causal and accommodating the necessary delay caused by the shift elsewhere in the system.

3.2.3 Linearity

The linearity of the mixing is a major factor affecting the difficulty of the separation. While the simplest case to consider is that of a linear mapping, under which assumption much of the earlier research of BSS was performed, two other active areas of research exist :

- *non-linear* the most general case (the linear case can be considered a subset of this)

- *post non-linear* a two-stage transformation, consisting of a linear mixing followed by a non-linear mapping

The principal reason for the difficulty of the non-linear case is in determining the inverse mapping — often already complicated for convolutive mixtures. Only the linear case is considered in this thesis.

3.2.4 Generality

A common distinction between solutions is the range of signal types to which the solution can be applied. Rather than being a direct consequence of the signal type, this limitation is usually due to the characteristics of the signal. As well as being classified as stationary or non-stationary, a signal can also be assessed for other properties, according to values of its fourth-order moment (kurtosis). This classification measures the 'peakedness' of the signal and thus describes how closely the probability density function of the signal resembles that of a Gaussian distribution. This results in the signal being categorised as super-Gaussian, Gaussian or sub-Gaussian (leptokurtic, mesokurtic and platykurtic respectively) [55]. Speech, for example, has a high kurtotic value, as its distribution is far more peaked than that of a Gaussian.

Some algorithms will only work for signals that exhibit particular characteristics, *e.g.* those that are super-Gaussian. Where this sort of limitation exists, it is often to allow a simplification in the learning rule, such as defining the direction in which to search for the maximum or minimum of the objective function. This simplification usually confers accelerated convergence or higher performance than the more general algorithms, which are capable of achieving a more modest performance over a much wider range of signals.

Consequently, advanced knowledge of the characteristics of the signals to be processed can be used in selecting an algorithm that offers optimum performance for that type of signal.

3.2.5 Complexity

Although many of the algorithms used to solve blind separation and deconvolution problems are mathematically complicated (as mentioned previously in Section 3.1.4) relatively simple update rules do exist, at least for instantaneous solutions. These include those used by Jutten &

Hérault [56] and Hyvärinen & Oja [57].

These update rules are usually easier to implement than the intricate matrix manipulation required by Amari *et al.*'s natural gradient approach [58]. Although the more complex techniques normally yield superior separating performance in terms of speed and degree of separation achieved as well as stability and reliability, low-complexity methods have also been developed to combat these problems. Low-complexity rules typically make use of either a large number of computationally inexpensive operations, or a smaller number of more powerful operations. Fiori has compared the computational cost and performance of various low-complexity signal separation algorithms.

As well as their ease of implementation in hardware, or directly in silicon, low-complexity approaches are often also more biologically plausible, thereby helping to explain a possible mechanism for how the cocktail party problem may be solved by the brain.

3.3 Assumptions and considerations

There are a few points pertaining to the assumptions made in modelling the problems that are worth noting before the discussion of existing solutions. These concern ideals or conditions adopted to simplify the modelling that are not realistic, and may, in fact, mask other real-world characteristics.

First is the issue of the number of sources (n) versus the number of sensors (m). There should ideally be at least as many sensors as sources ($n \leq m$), otherwise no exact solution is possible. If there are fewer sensors than sources ($n > m$), BSS may be carried out on the incoming mixture and the m most significant dimensions dealt with, where m is the number of sensors available. The m most energetic signals are thus separated, with the remaining signals acting as noise, and degrading the quality of the separated signals. Some techniques that attempt to further separate the remaining sources now also exist. If there are fewer sources than sensors ($n < m$), then it will be possible to separate all of the sources, subject to other limiting constraints, such as the mixing being invertible. Under certain conditions the remaining sensor outputs will be undefined, or possibly tend to zero mean Gaussian noise [56]. In this thesis, it should be assumed that $n = m$, unless otherwise stated.

In the work on convolutive mixtures, the inverse filters can be implemented using either

Finite Impulse Response (FIR) filters or Infinite Impulse Response (IIR) filters. FIRs must be long enough to give a 'reasonable' approximation of the desired inverse. 'Reasonable' approximations may be achieved using a filter length of two to four times that of the original. IIRs offer the advantage of requiring far fewer taps to approximate the inverse, but cannot be used if the inverse filter is unstable or non-causal. In either case, the length of the original filters is unknown in a truly blind scenario. This suggests making the weight set long enough to accommodate a range of potential filter lengths, regardless of its architecture. It should be borne in mind, though, that the computational effort of the algorithms increases with the length of the separating filters (as there are more tap weights to find). In an attempt to avoid unnecessary processing, the configurations used in the experiments in this thesis will use a weight set of length twice that of the filters used to create the input data.

Since the algorithms developed aim to be as general as possible, it is undesirable to limit the types of filters that can be dealt with. One characteristic of the filter than must be considered is whether or not it is minimum-phase. Non-minimum-phase filters have poles or zeros outside the unit circle, and consequently, their inverses are non-causal [59]. As has previously been stated, in Section 3.2.2, non-causal filters can be made causal by the introduction of the necessary delay. This has the effect of shifting the taps of the filter and requires buffering of the signals to accommodate this. As a result, care must be taken in the selection of the algorithm and architecture to be used in implementing a solution to a combined separation and deconvolution problem, or the system may be incapable of solving the intended problem.

Another constraint imposed by some of the algorithms relates to the stationarity and Gaussianity of the source signals. For stationary source data to be separable, at most one of the signals can be Gaussian. This is due to the fact that by the Central Limit Theorem [1], a mixture of more than one Gaussian random variable cannot be distinguished from a single random variable with a mean and variance given by the sum of the individual random variables. Likewise, if all of the sources are non-stationary, then at most one of them can be non-Gaussian for similar reasons of indeterminacy [40,52].

The majority of the methods described in the remainder of this chapter assume that there is no noise added to the mixtures of the signals at any stage of the processing. If there is noise, then unless the magnitude of any noise considered is large compared to that of the other signals being processed, it can often be ignored. However, in cases where the noise must be considered, it can complicate the separation process, particularly if the form or distribution of

the noise is not known in advance. Consideration of noise is an issue that was frequently and deliberately ignored in the earlier work in this field, but that is now receiving more attention as it is recognised as being an important (albeit difficult) constituent of the real world problems of blind separation and deconvolution.

3.4 Approaches to blind source separation

This section reviews some of the principal techniques and approaches used in solving the blind source separation problem. It does not attempt to be exhaustive, but instead highlights noteworthy solutions, and those pertinent to this thesis. Approaches to solving the combined blind separation and deconvolution problem appear in Section 3.5.

The approaches that follow have been categorised into two groups : those that are based on a mathematical or signal processing oriented perspective, and those that are based more on a neural network framework. Both groups, however, share some common aspects.

3.4.1 Common aspects

Despite the distinctions of these two groups of solutions, since they are both attempting to solve the same problem, it is only natural that there are certain common elements or ideas. The following concepts are among the more important of these ideas and are used, in one form or another, in several of the approaches described later.

3.4.1.1 Equivariance

The method of Equivariant Adaptive Separation via Independence (EASI) [60] is one of a number of related variations on the gradient descent approach described previously, that defines the update in terms of the global system, characterised by a matrix C — the product of the mixing and the unmixing matrices :

$$\mathbf{C} = \mathbf{WA} \tag{3.7}$$

In order to be described as equivariant, the method for estimating a matrix $\mathbf{W}$ based on a vector $\mathbf{x}$ must satisfy the following criterion :

$$\mathbf{M}(\mathbf{Wx}) = \mathbf{WM}(\mathbf{x}) \tag{3.8}$$

for any invertible matrix $\mathbf{M}$ of the same dimensions as $\mathbf{W}$. This means that the update rule can now be defined as :

$$\mathbf{C}(t+1) = \mathbf{C}(t) + \eta(t)\mathbf{G}\left(\mathbf{C}(t)\mathbf{u}(t)\right)\mathbf{C}(t) \tag{3.9}$$

from [18]. In this way, the performance of the algorithm is determined by the matrix $\mathbf{C}$ and the source distributions, and not by the particular values of $\mathbf{A}$ and $\mathbf{W}$ at any instant in time. The matrix $\mathbf{C}$ should ideally converge to the permutation matrix for the system.

In order to maintain uniform performance across all mixing matrices, the objective function to be maximised or minimised should be relative to the distributions of the sources, rather than having some absolute maximum or minimum. This gives optimum performance, *i.e.* updates in the direction of the steepest gradient of the objective function. The update rule derived from this *relative gradient* approach is :

$$\mathbf{W}(t+1) = \mathbf{W} - \eta\mathbf{f}'\left(\mathbf{y}(t)\right)\mathbf{y}^T(t)\mathbf{W}(t) \tag{3.10}$$

where $\mathbf{f}'(\mathbf{y})$ is the gradient of $f()$ at $\mathbf{y}$. $\mathbf{f}'()$ is sometimes referred to as the 'score' function, in the context of Maximum Likelihood approaches (see Section 3.4.2.3).

A similar method was developed independently by Amari, Cichocki and Yang [58], under the name of the *natural gradient* approach, following an information geometry perspective. This method discussed tuning the gradient descent for situations where the geometric bases are not mutually orthogonal, and was based on earlier work by Cichocki *et al.* [61,62]. This approach has proven extremely popular due to the fast convergence it affords.

These methods are computationally inexpensive, but offer inherent gain control on the outputs, limiting them to unity. They are also very tolerant of the choice of non-linearities in the learning rule, and generally have better convergence characteristics than the original algorithms.

3.4.1.2 Entropy Estimation

A clear choice for an objective function for the BSS / ICA problem is the mutual information between any pair of the outputs, $I(y_i, y_j)$, since this should tend to zero as the separation proceeds. It has been noted by Haykin [18], amongst others, that minimising the mutual information between output components is equivalent to minimising the Kullback-Leibler divergence between the joint probability density function $p_{\mathbf{Y}}(\mathbf{y}, \mathbf{W})$, parameterised by the weight set $\mathbf{W}$, and the product of the marginal probability density functions given by :

$$\widetilde{p_{\mathbf{Y}}}(\mathbf{y}, \mathbf{W}) = \prod_{i=1}^{m} \widetilde{p_{Y_i}}(y_i, \mathbf{W}) \tag{3.11}$$

The divergence is often expressed in terms of the entropies of these distributions :

$$D_{p||\tilde{p}}(\mathbf{W}) = -h(\mathbf{y}) + \sum_{i=1}^{m} \tilde{h}\left(y_i\right) \tag{3.12}$$

which makes the link to the mutual information constraint more clear, but requires an estimate of the marginal entropies $\tilde{h}(y_i)$ which can only be approximated using a series expansion of the higher order cumulants of y_i.

Negentropy has also been used as an objective function, and is defined as the difference in entropy between an arbitrary distribution, p_y, and the equivalent entropy of a Gaussian distribution, p_G, of equal mean and covariance :

$$J\left(p_u\right) = H\left(p_G\right) - H\left(p_y\right) \tag{3.13}$$

The sum of the kurtoses of the outputs, and combinations of other higher order moments and cumulants have also been proposed as contrast functions elsewhere [63–67]. It is generally accepted that using up to the fourth order term gives sufficient accuracy in most circumstances, although Bell & Sejnowski [10] argued that all of the higher terms should strictly be considered.

The update rules for the weight set, based on these objective functions, will be of the form :

$$\Delta\mathbf{W} = \eta\left(\mathbf{W}^{-T} - \varphi(\mathbf{y})\mathbf{x}^T\right) \tag{3.14}$$

$$= \eta\left[\mathbf{I} - \varphi(\mathbf{y})\mathbf{y}^T\right]\mathbf{W} \tag{3.15}$$

where $\varphi(y_i)$ is a non-monotonic activation function used to introduce HOS into the update

mechanism. These activation functions are frequently non-linear and odd, and good results have been achieved using the logistic function :

$$\varphi(y_i) = \frac{1}{(1 + e^{-y_i})} \tag{3.16}$$

and with $\varphi(y_i) = y_i^3$ or $\varphi(y_i) = \tan^{-1}(y_i)$.

A point made by Amari, Cichocki and Yang in [58], is concerned with the numerical expansion used in the estimation of the marginal entropies, as part of the measurement of the dependency between the independent output components. This is required as the dependency is usually measured by the Kullback-Leibler divergence between the joint distribution and the product of the marginal distributions, which incorporates the marginal entropy. The expansion originally used by Comon [68] and Bell & Sejnowski [10] was the Edgeworth expansion. Amari *et al.* [58], however, use the Gram-Charlier expansion instead, as it does not lose the fourth order term in the standard truncation used, unlike the Edgeworth expansion. They showed that this term is of great importance in the situations considered.

3.4.1.3 Deflation

The deflation approach proposed by Delfosse and Loubaton [69] allows the extraction of a single component of the mixed sources at a time, rather than attempting to simultaneously extract them all in parallel. With a parallel approach, convergence to a stable solution cannot always be guaranteed. In contrast, the authors demonstrated that by separating out only one source at a time, the convergence property of the algorithm would hold for any number of sources.

The deflation approach means that the full number of signals need not be known in advance, and not all separated out if this is not desired. It is similar in some respects to the construction of a network of units using Oja's PCA rules, as each of them learns to filter out a single component of the mixture. Fyfe & Girolami [70] have also adopted a deflationary approach in some of their work with Exploratory Projection Pursuit networks (see Section 3.4.3.5), and show that this outperforms a parallel network in their experiments.

3.4.1.4 Sphering

For situations where computational load is not an issue, the input data can be *sphered*, or *pre-whitened*, prior to being passed to the separating network [40]. Sphering whitens the inputs, resulting in a set of n mutually uncorrelated, unit variance signals. This greatly eases the task of the separation, simplifying it to finding an appropriate rotation to convert the mutually uncorrelated signals into statistically independent ones.

Sphering, however, also has the undesirable effect of flattening a speech signal, distorting the way it sounds and making it most unnatural. In applications that are to operate on speech signals, sphering is often employed in the training stage of the system, and the resulting weight set applied to the original signal to yield outputs that are as undistorted as possible.

3.4.2 Non-neural approaches

The set of topics considered here have been termed "non-neural" approaches as they are not explicitly defined around a neural network architecture, although they still make use of adaptive processing. Rather, they exploit mathematical properties and signal processing techniques, and apply them to the blind separation problem. This is not an exhaustive list of such approaches — but a number which highlight some interesting considerations or perspectives.

3.4.2.1 Independent Component Analysis

Independent Component Analysis (ICA) is a method for investigating structure within a multivariate data set, first described by Comon in [68]. It is related to the well-known approach of Principal Component Analysis (PCA). By projecting the data onto a *feature space* of a different dimension, any inherent organisation of the data may become apparent. While PCA seeks orthogonal components, they are only decorrelated — independent up to second order. Independent Component Analysis (ICA) extends this by ensuring that the components are statistically independent, and that they are no longer constrained to be orthogonal. This is because the ICA constraint is more strict than the decorrelation constraint for PCA — the mutual information measure takes account of all the higher order statistics (HOS) of the input signals, whilst the PCA one (pairwise decorrelation : $\langle y_i y_j \rangle = 0$) depends only on the second order statistics [26].

The aim of ICA is to be able to recover the original data when presented with only the transformed representations of the feature space. It is often also desirable to be able to determine the transformation back from the feature space. The mapping from the data space to the feature space can be considered a mixing of the data, as this is what is effectively done by the projection of the data vector along each of the directions of the independent component vectors. The corresponding reverse transformation is consequently the recovery of the original data, by unmixing or separation. This can be achieved only up to permutation and scaling of the data — that is to say that the order in which the original data are retrieved cannot be guaranteed, nor can their relative magnitudes. However, since it is usually the *shape* of the waveforms that is of interest, this is acceptable.

3.4.2.2 JADE

The Joint Approximate Diagonalization of Eigen-Matrices (JADE) is a technique that was originally developed for blind identification and blind beamforming work, by Cardoso and Souloumiac [71]. It works by optimising a contrast function based on the diagonalisation of the eigenmatrices — that is, it attempts to find a single matrix that maximises the contrast function over the whole set of the matrices. The eigenmatrix decomposition of the resulting *joint diagonaliser* provides the solution to the ICA problem.

Some of the key benefits of JADE are that it works without any parameter tuning and has been successfully employed in this manner to a number of different applications in a variety of fields.

3.4.2.3 Maximum likelihood

Maximum Likelihood (ML) is a method of statistical estimation which determines the most plausible distribution from which a particular sample will have been drawn [18]. It is generally applicable to models that have latent variables — that is, where the observable values of a problem are drawn from some underlying (possibly mixed) probability distributions, such as those in Hidden Markov Models (HMMs) [72]. This technique can be used to define a contrast for solving problems using gradient-based approaches, as the value of the likelihood can be maximised.

ML has been applied to the ICA / BSS problem, and shown in [18], [73] and [72] to be equivalent to the information maximisation approach of Bell and Sejnowski [10] (and see also

Section 3.5.1.1), when the source distributions are considered the latent variables. Maximising the log-likelihood functions for the outputs, parameterised on the weight set $\mathbf{W}$ is again shown to be equivalent to minimising the Kullback-Leibler divergence between the output distribution and the hypothesised source distributions.

MacKay [72] further extends this by developing a covariant algorithm, which is quite simple and yet converges rapidly. Several other, more recent papers, applying ML to ICA problems can be found in [74].

3.4.2.4 Nonholonomic orthogonal learning constraints

Amari, Chen and Cichocki [54] have formalised the difficulties encountered by learning algorithms when dealing with non-stationary signals. They show that the constraints imposed by some learning algorithms on the magnitude of the recovered signals, to help resolve the unknown scaling factors, effectively tie the separating matrix to the sources (via the inputs). Consequently, if the magnitude of the sources changes rapidly, so will the magnitude of the separating matrix. This may lead to instability. The authors proposed a new, nonholonomic learning algorithm which used different constraints which do not result in this tying, and consequently enable the algorithm to perform well even with rapid changes in source magnitude.

The generalised learning algorithm is given by :

$$\mathbf{W}(t+1) = \mathbf{W}(t) - \eta(t)\mathbf{F}\left\{\mathbf{y}(t)\right\}\mathbf{W}(t) \tag{3.17}$$

where the entries of the matrix $\mathbf{F}(\mathbf{y})$ are defined as :

$$f_{ij} = \delta_{ij}\lambda_{ij} + \alpha_1 y_i y_j + \alpha_2 \varphi_i\left(y_i\right) y_j - \alpha_3 y_i \varphi_i\left(y_j\right) \tag{3.18}$$

δ_{ij} is the Kronecker delta, $\lambda_{ii} = -\alpha_1 y_i^2 - \alpha_2 \varphi_i\left(y_i\right) y_i - \alpha_3 y_i \varphi_i\left(y_i\right)$, $\varphi_i\left(y_i\right) = -\frac{\dot{p}_i(y_i)}{p_i(y_i)}$ and α_1, α_2 and α_3 are adaptively determinable parameters.

The approach by Amari, Chen and Cichocki [54] considers all possible separating solutions that differ only in the scale of the components — an equivalence class, the set of all equivalence classes making up the entire solution space. By allowing only updates that are orthogonal to a particular solution's equivalence class, the convergence trajectory of the weight set is

guaranteed not to include redundant components that could otherwise lead to instability. They demonstrate the local stability of their methods around the desired separating solution for mixtures of signals of the same kurtotic sign.

3.4.3 Neural network approaches

A relatively large proportion of the approaches to solving BSS problems are set on a neural network basis. This neural link is appropriate and appealing since the human brain is clearly capable of solving the cocktail party problem (Section 3.1.1.1). Consequently, it seems reasonable to postulate that it should be possible to construct an artificial implementation that achieves the same end result. Some of these implementations are more biologically plausible than others — *e.g.* operating locally without requiring global knowledge of the whole network, a desirable property for this field — but this is not an essential attribute of the solutions. Different frameworks, even within the neural network foundation, have lead to several distinct approaches to the problem, the main ones of which are presented below.

3.4.3.1 Hérault and Jutten

Hérault, Jutten and Ans were the first to propose a neuromimetic solution [75] to the problem of Blind Source Separation. This was re-addressed in the classic 1991 paper, [56], where they set out the network architecture and learning rules to be used. The recursive network structure, that subtracts each of the current source estimates from each of the mixtures, is shown below in Figure 3.2 (based on [56]).

Mathematically, this can be represented as :

$$\mathbf{y} = \mathbf{x} - \mathbf{W}\mathbf{y} \tag{3.19}$$

which is equivalent to :

$$\mathbf{y} = (\mathbf{I} + \mathbf{W})^{-1}\mathbf{x} \tag{3.20}$$

Using these foundations, the square of the i^{th} output, y_i^2 can be considered an error term, which can be used as the objective function for the system. Gradient descent can then be applied to

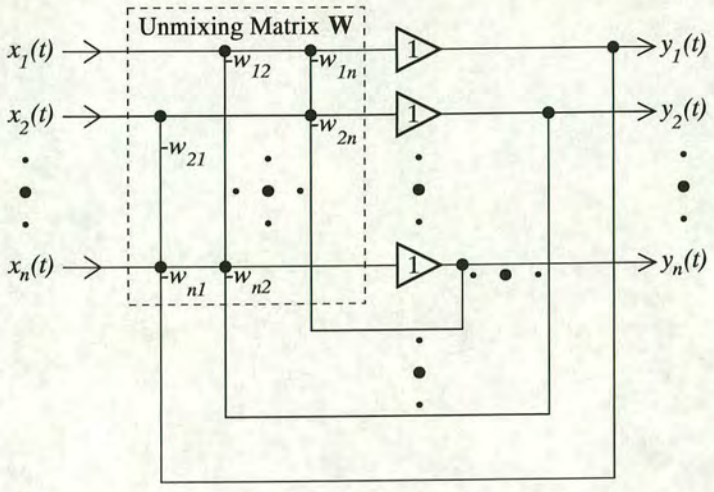


Figure 3.2: *The Hérault-Jutten network*

this, in terms of the weight set **W** to give :

$$\frac{dw_{ik}}{dt} = \eta f\left(y_i(t)\right) g\left(y_k(t)\right) \tag{3.21}$$

where $f()$ and $g()$ are two different, odd, non-linear functions such as $f(x) = x^3$ and $g(x) = \tan^{-1}(x)$ used to introduce higher-order moments as part of an independence test. This method is quite stable for a reasonable range of functions, many of which are acceptable under the biological plausibility constraint, aided by the fact that the rules proposed are local.

Hérault and Jutten then go on to explain the principles of independent component analysis, comparing it with principal component analysis, and how it relates to their new algorithm. Mathematical and statistical analyses of their work have been presented by the authors themselves [76] and by others [77–79], and details of hardware implementations of, and experiments with, their networks can be found in [37,38].

3.4.3.2 Information Maximisation

Bell and Sejnowski's classic paper [10] sets out a new framework for the BSS / ICA problem, casting it into an information theoretic perspective. Their approach was based on work by Linsker [11] whose infomax principle had suggested that unsupervised networks with relatively simple learning rules could learn to maximise the information transferred between their inputs and their outputs, whilst at the same time, minimising the mutual information between their

outputs. Maximising the information transferred between the inputs and the outputs ensured that as little signal degradation as possible occurred. The mutual information between the outputs is the result of the mixing, and hence minimising it has the effect of separating the signals.

The method developed was a generalisation of Linsker's infomax principle to non-linear units. When extended to a whole network, this maximises the individual output variances, and reduces the redundancy in the output layer. The information transferred between input and output layers of a network is defined by :

$$I(\mathbf{y},\mathbf{x}) = H(\mathbf{y}) - H(\mathbf{y} \mid \mathbf{x}) \tag{3.22}$$

(which holds even for a single layer network). Since entropy is relative in the continuous case, the gradients of these information theoretic quantities were considered with respect to the weights of the network, in determining an objective function for this part of the constraint. The learning rule based on this simplifies to the following :

$$\frac{\delta}{\delta w}(I(\mathbf{y},\mathbf{x})) = \frac{\delta}{\delta w}(H(\mathbf{y})) \tag{3.23}$$

as $H(\mathbf{x})$ does not depend on w. Consequently, the mutual information communicated between inputs and outputs can be maximised by maximising $H(\mathbf{y})$ alone.

When processing the inputs using a non-linear function, $g(y)$, such as the logistic function (Equation 3.16) to generate an output, z, maximum information preservation is achieved when the sloping part of the sigmoid is optimally aligned with the highest density parts of the inputs, as shown in Figure 3.3 (based on [10]).

$$y = wx + w_0 \tag{3.24}$$

$$z = \frac{1}{1+\exp^{-y}} \tag{3.25}$$

$$\Delta w_0 \propto 1 - 2z \tag{3.26}$$

$$\Delta w \propto \frac{1}{w} + x(1-2z) \tag{3.27}$$

The learning rules associated with this point, are given for the single neuron case as Equations 3.26 and 3.27. They match the steepest part of the curve to the highest input density, and spread the slope of the sigmoid curve to match the variance of $f_x(x)$. The Δw rule is

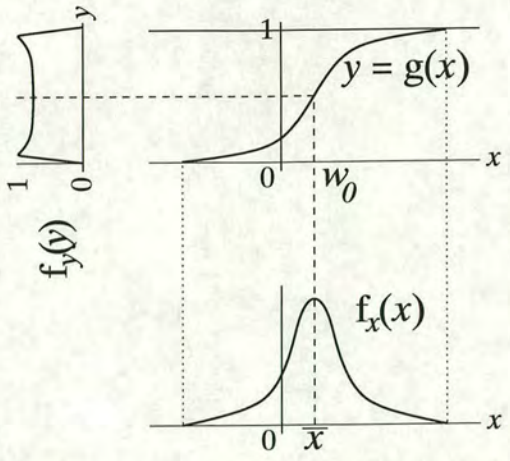


Figure 3.3: *PDF Matching and Learning Rules*

anti-Hebbian, which keeps z away from the uninformative situation where it is saturated at either zero or one. This alone would eventually force the weights to zero, so an anti-decay term is used to keep the weights from becoming too small.

The PDF matching results in better performance in the adaptive search for the unmixing matrix $\mathbf{W}$, by reducing the likelihood that the network will become stuck in a local minimum of the mutual information search space. These minima are features of the surface defined by the objective function, based on the mutual information from the output (estimated source) distributions.

When extended to an n-node network, the inputs, outputs and bias terms become vectors, and the weight set becomes a matrix. The system can then be illustrated as shown in Figure 3.4.

The expanded learning rules are summarised as follows :

$$\mathbf{y} = \mathbf{W}\mathbf{x} + \mathbf{w}_0 \tag{3.28}$$

$$\mathbf{z} = g(\mathbf{y}) \tag{3.29}$$

$$\Delta\mathbf{w}_0 \propto \mathbf{1} - \mathbf{2z} \tag{3.30}$$

$$\Delta\mathbf{W} \propto \left[\mathbf{W}^T\right]^{-1} + (\mathbf{1} - \mathbf{2z})\mathbf{x}^T \tag{3.31}$$

and for individual weight, w_{ij}, this last rule becomes

$$\Delta w_{ij} \propto \frac{\mathrm{cof} w_{ij}}{\det \mathbf{W}} + x_j\,(1 - 2z_i) \tag{3.32}$$

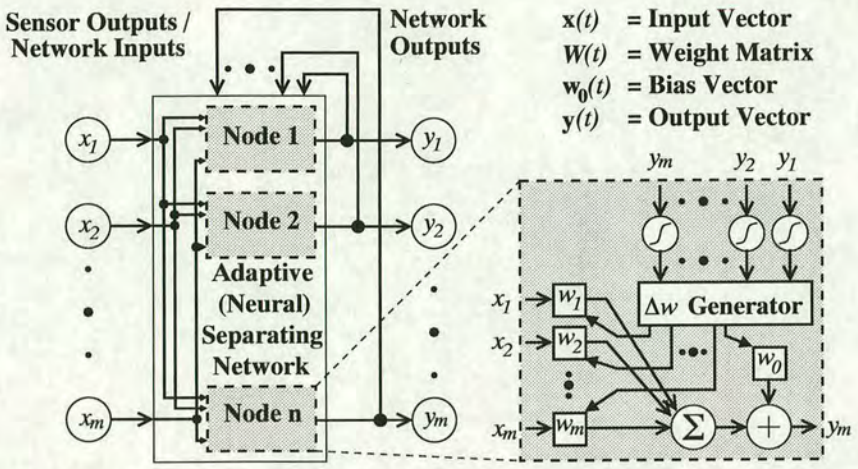


Figure 3.4: *The single layer infomax network*

These last two rules are computationally very expensive, and undesirable for efficient implementations. They also break the biological plausibility ideology as they are non-local — requiring global knowledge of the rest of the network.

However, in its favour, the approach taken can be extended to deal with causal filters, weights with time delays, generalised sigmoid functions, and combinations of these complications. It allows the same ideas to be used in solving blind deconvolution, where statistical dependencies across time must be removed — a process known as whitening [10].

Overall, the system performs very well, particularly when compared to the alternatives available at the time of its development. It renewed interest in the field and formed the foundation for a number of other pieces of work, most noticeable, perhaps, the application of these techniques to electro-encephalographic (EEG) recordings [26, 80].

3.4.3.3 Single Unit Learning Rules

Hyvärinen and Oja [81] have developed sets of learning rules and contrast functions based on kurtosis maximisation or minimisation that enable networks to be constructed from individual neurons each of which is capable of extracting a single independent component from a mixture. They do not require any global knowledge of the rest of the network, which offers a distinct advantage in terms of modular design, particularly for hardware or firmware implementations since it eases the otherwise critical constraint of needing to know the size of the network in advance. The ideas are related to the deflationary approaches discussed previously, but make a

number of assumptions concerning the bounds on the signal and network variables.

A number of rules are presented, for dealing with mixtures of signals with known kurtotic sign, and for sphered and non-sphered data. The general form of the objective function is given by :

$$J(\mathbf{w}) = \alpha E\left\{\left(\mathbf{w}^T\mathbf{x}\right)^4\right\} + \beta F\left(E\left\{\left(\mathbf{w}^T\mathbf{x}\right)^2\right\}\right) \tag{3.33}$$

where $\alpha, \beta > 0$ are arbitrary scaling functions and F is a penalty function designed to take account of the variance of the estimated sources $\mathbf{W}^T\mathbf{x}$ not being unity [81]. $J()$ must be maximised or minimised for mixtures of positive or negative kurtoses respectively. If the kurtoses of the independent components are not known in advance, two units must be used to separate out each component — the second providing an estimate of the kurtosis of the output of the first unit [57].

When extending this to deal with multiple independent components, a feedback term must be included from each of the neurons that forces the solution away from the other, already found, independent components. Such networks can be constructed as a layer of single units, if the sign of the kurtosis of the signals to be extracted can be determined in advance, otherwise the layer must be composed of the pairs of units described above.

The fast, fixed-point algorithm [82] proposed by the same authors can be used to further speed up convergence of these or other BSS / ICA solutions. This simple algorithm, which requires only the input data to be sphered and an on-line estimate of the kurtosis, allows the extraction of signals of either sign of kurtosis, in a hierarchical manner and can be used in either a batch or a semi-adaptive system. Its cubic rate of convergence is much faster than most of the other ICA-solving algorithms.

3.4.3.4 Non-linear PCA

The simple Hebbian-based PCA networks developed by Oja (see Haykin [18] for a summary of this work) can be configured to use non-linear activation functions, enabling them to perform ICA on a pre-whitened input vector, $\mathbf{v}$. The learning rule is similar to that presented earlier for the generalised ICA update [83] :

$$\Delta\mathbf{W} = \eta[\mathbf{v} - \mathbf{W}\mathbf{g}(\mathbf{y})]\,\mathbf{g}(\mathbf{y}^T) \tag{3.34}$$

Without the pre-whitening step, the range of signals which it can separate successfully is considerably diminished [40]. However, the main advantage of this technique, is that it can be implemented as a simple, single-layer, locally governed ANN.

In [83], the authors simulate a recursive non-linear PCA network that uses a fast implementation of the least-squares algorithm in its update rule. This converges up to an order of magnitude more quickly than the normal non-linear PCA rules. Oja, in [84], shows the equivalence of this technique with that of the Maximum Likelihood approach [73], [72], and consequently the information-maximisation criteria, too.

3.4.3.5 Exploratory Projection Pursuit

Exploratory Projection Pursuit (EPP) is a technique that can be used for discovering structure in high dimensional datasets. The data are projected onto a low dimensional subspace, and indices defined that quantify how "interesting" the projection is. The direction that gives the most interesting projection can then be explored for structure within the data.

Girolami and Fyfe have applied this technique to ICA by using negentropy and kurtosis based indices [85] in a EPP network. Maximisation of these indices corresponded to the direction furthest from Gaussianity — which is what is desired in most stationary mixing cases.

Kurtosis-based contrasts and their limitations have been discussed previously, and similar considerations apply here in their use as projection pursuit indices, including the possible requirement for pre-whitened data depending on the non-linearity used.

The original network proposed by Girolami and Fyfe [86] was a two layer structure with feedback, but no lateral connections. The learning rule used is of the form :

$$\Delta w_{ji} = \eta_t f\left(\sum_{j=1}^{N} w_{ij}x_j(t)\right)\left\{x_j(t) - \sum_{k=1}^{M} w_{kj}\sum_{j=1}^{N} w_{ij}x_j(t)\right\} \tag{3.35}$$

where $f()$ is a non-linear function. The authors have also proposed variations on this, including a laterally inhibited deflationary network [85] and an Extended Exploratory Projection Pursuit network capable of tracking non-stationary inputs [70]. They demonstrated that lateral connections aid convergence, that the deflationary approach performs better than a parallel network, and that self-connections are essential in finding the solution. They further

commented that using random sampling, rather than consecutive sampling, aids convergence in some cases.

3.4.3.6 Time dependent learning rules

Matsuoka *et al.* [51], [52] approached the issue of the blind separation of non-stationary signals by defining the cost function used in the network's adaptation rule to be time dependent. This achieves its minimum value when the network outputs are uncorrelated with one another, and directly incorporates the signal non-stationarities into the problem, rather than transforming the inputs to change the situation, as do some of the other algorithms. The network has a recurrent structure, with inhibitory lateral connections but no self connections (*i.e.* $w_{ii} = 0$), and the learning rule is of the form :

$$\eta \frac{d\mathbf{W}}{dt} \doteq \left(\mathbf{I} + \mathbf{W}^T\right)^{-1} \left\{ \left(\text{diag}\Phi(t)\right)^{-1} \mathbf{y}(t)\mathbf{y}(t)^T - \mathbf{I} \right\} \tag{3.36}$$

where $\mathbf{I}$ is the identity matrix, and $\Phi(t) = \text{diag}\left\{\phi_1(t), \ldots, \phi_N(t)\right\}$, $\phi_i(t) = \left\langle y_i^2 \right\rangle$ and

$$\eta' \frac{d\phi_i(t)}{dt} = \phi_i(t) = y_i^2(t) \qquad (i = 1, \ldots, N) \tag{3.37}$$

They recently extended this work to deal with convolutive mixtures [87].

Recently, Fyfe and Cichocki [88] demonstrated that Matsuoka *et al.*'s original algorithm is not linear as originally claimed. They then presented a hierarchical model and a parallel model, both capable of separating out non-stationary independent components from linear mixtures.

3.4.3.7 Pre-filtering

Barros and Ohnishi [53] addressed the problem of non-stationary signals by applying a low-pass filter to each of the inputs, thereby reducing the effect of rapid transients in the signal amplitude. The transformation of the input vector $\mathbf{x}$ can be described by :

$$\mathbf{x}' = \frac{\mathbf{x}}{\Gamma\left[\sum_i {x_i}^2\right] + c} \tag{3.38}$$

where $\Gamma(\cdot)$ is a low-pass filter operator, and c is a small constant to avoid division by zero. Having normalised the weights, the updates for the weight set were then calculated using :

$$\Delta \mathbf{W} = \eta \left(\mathbf{I} - 2\mathbf{u}\mathbf{x}'^{\mathbf{T}} W^T \right) \mathbf{W} \tag{3.39}$$

$$\Delta \mathbf{w_0} = \widetilde{\eta} \left(-2\mathbf{u} \right) \tag{3.40}$$

with $\mathbf{u} = \tanh\left(\mathbf{B}\mathbf{x} + \mathbf{b_0}\right)$ and $\eta_{k+1} = \eta_k - \eta_k^2$.

They also showed that this pre-filtering flattened the inputs in the time domain, leading to better separating performance.

The pre-filtering works by slowing down the rate of change of variance in the input signals, thereby reducing the potential for numerical instability in the convergence of the network's weight set to the desired solution. Low frequency trends in the signals, such as possible drift in the signal mean, are not filtered out, but these have a much less drastic effect on the convergence. It should be noted that such pre-filtering affects only the generation of the updates for the network, and does not transform its outputs, so the signals retrieved are not distorted by the filtering.

3.4.4 Summary of approaches

The key aspects of the approaches described have been summarised in Table 3.2 for ease of reference.

3.5 Approaches to blind separation and deconvolution

Rather than classify the approaches to the combined blind separation and deconvolution problem into "neural" and "non-neural" groupings, they are more conveniently grouped by a different aspect of their processing — their signal representation domain. These two categories are Time Domain processing, and Frequency Domain processing.

The time domain is that in which most of the processing considered so far has been carried out. The representations of signals show a varying amplitude for successive instances in time, the amplitude conveying the information of the signal's value at that instant.

In the frequency domain, the signal is viewed as being composed of the sum of a number of

	Method	Key aspects
Non-neural	Independent Component Analysis	Powerful mathematical framework
	JADE	No parameter tuning required
	Maximum Likelihood	Allows use of prior knowledge
	Non-holonomic Learning Constraints	Reduces instability due to signal non-stationarity
Neural	Hérault & Jutten	Original neural solution
	Information Maximisation	Classic entropy-based framework
	Single Unit Learning Rules	Simple, modular design
	Non-linear PCA	Biologically plausible learning rules
	Exploratory Projection Pursuit	Extendable for deflationary approach, and tracking non-stationarities
	Time Dependent Learning Rules	Implicitly deals with signal non-stationarity
	Pre-filtering	Simple modification to Infomax approach

Table 3.2: *Summary of key aspects of Approaches to Blind Separation*

different components from the frequency spectrum. Each of these components is a sine wave of a different frequency, and the proportion of each frequency to be summed is given by the amplitude of that component.

The time and frequency domains are orthogonal to one another. The frequency domain can be reached from the time domain by carrying out a Fast Fourier Transform (FFT) on a signal. An Inverse Fast Fourier Transform (IFFT) can be used to achieve the reverse mapping back to the time domain. Some approaches involve processing in both domains. Where an approach considers both domains, this will be noted accordingly.

3.5.1 Time domain approaches

Much of the initial work on convolutive mixtures focused on extending the processing carried out on instantaneous mixtures to multiple time points, one for each tap of the inverse filters to be estimated. As the solutions to the instantaneous problem were largely based in the time domain, so were many of the early solutions to the convolutive problem.

3.5.1.1 Information maximisation

As mentioned previously in Section 3.4.3.2, Bell & Sejnowski's information maximisation approach can be extended to deal with the combined separation and deconvolution problem.

Torkkola [5] extended Bell & Sejnowski's work on separation and deconvolution to further permit generalised delays in the signals, rather than just a specific delay associated with a particular weight. This makes the model more realistic, as the propagation of any signal takes a finite amount of time. Hence signals from sources that are not equidistant from a sensor will arrive at different times, resulting in a relative delay between them. Torkkola noted that this is related to the multipath problem, and modified Bell & Sejnowski's model to incorporate adaptive delays, as proposed by Platt and Fagin [89]. After deriving the local adaptation rules required to learn the variable delays, some experiments were undertaken on speech signals.

It was noted that the delays were learned before the weights converged to the target values of the inverse FIR, and that if the delays converged to the wrong values, it was unlikely that the weights would converge to a separating solution. Consequently, initialising the delays to appropriate values is crucial, due to the periodic nature of speech which can otherwise attract the delays to correlations within the signals. This problem can be overcome by whitening the signals over a short period of time (2–3 ms) [5], and learning the delays and weights from these pre-processed signals. These can then be applied to the original signals to recover the desired outputs.

This work was further extended in [3] to deal with convolved signals. However, the framework developed was only capable of dealing with multiple non-interfering sources, without introducing whitening effects. The delays to the first non-zero tap of the filter can also be learned, allowing much shorter separating filters to be employed, which results in less taps to be learned. This can also be achieved by using IIR filters instead of FIRs, providing that the inverse filter is stable. A methodology for deriving the required learning rules to support these was developed in [2].

According to Lee, Bell & Lambert [90], Torkkola's work is limited by the absence of cross-connections at $t = 0$. Lee *et al.* present a full feedback system capable of solving any minimum-phase system, using a natural gradient algorithm [58], but note that it may suffer from local minima problems when processing periodic signals such as speech. They also note that real room responses are usually non-minimum-phase and develop frequency domain

learning rules for a feedforward architecture using FIRs that can deal with them. The learning rules are derived using Lambert's FIR polynomial matrix algebra [91], and hence move much of the processing to the frequency domain.

This FIR polynomial matrix algebra is also used to develop the learning rules in [59], which explains how the architecture uses the non-causal inverses generated by the algebra, by expressing them as the concatenation of a (causal) minimum-phase filter and an all-pass filter (delay). The authors also explain how the learning rules can be derived from a number of different perspectives such as information maximisation, maximum likelihood and negentropy. This solution is augmented by the authors in [92] by preprocessing the sources using the time-delayed decorrelation algorithm of Molgedey & Schuster [93]. This decorrelation algorithm gives an improvement in the convergence performance for cases where the sources are close to the sensors, but performs poorly when the sources are further away.

3.5.1.2 Natural Gradient Approach with Nonholonomic Constraint

Choi *et al.* [94] extend the natural gradient based multi-channel blind deconvolution solution proposed by Amari *et al.* [95] to incorporate the nonholonomic learning constraints of Amari, Chen & Cichocki [54] discussed in Section 3.4.2.4. They demonstrate the improved convergence of both a feedforward and a feedback system on binary data for the overdetermined case, where there are more sensors than sources.

3.5.1.3 Decorrelation

In [9], Nguyen Thi & Jutten propose several separation criteria based on the cancellation of fourth-order cumulants or moments and apply these to convolutive mixtures of various signals. They derive learning algorithms for these criteria and select one that avoids the possibility of spurious solutions. They also comment on the experimental efficacy of *non-permanent learning* in their experiments involving non-stationary signals, and in particular speech signals. Non-permanent learning avoids problems of varying signal energy in the outputs. It involves setting an energy threshold and only updating the filter coefficients that affect an output if the energy of that output is above the threshold.

Van Gerven discusses the importance of similar threshold setting techniques in [8] and considers the optimum settings for an "intermittent adaptation" threshold in conjunction

with the use of a Symmetric Adaptive Decorrelation architecture in [6]. This architecture is analysed by the authors in [96] and does not strictly require the use of these controlled adaptation techniques to overcome the otherwise detrimental effects of signal leakage on the update of other adaptive noise cancellers. They further point out that decorrelation alone cannot solve the separation of convolved sources if there is mixing at the first tap of the filter (zeroth-order). The algorithms developed require that the filters all be causal, and convergence is not guaranteed from arbitrary starting conditions.

In [87], Kawamoto *et al.* develop the learning rules proposed by Matsuoka, Ohya & Kawamoto [51] to deal with non-minimum-phase convolutive mixtures of non-stationary signals. The fact that the signals are independent, zero mean and non-stationary is the only information required about them. The inverse filters are estimated from the second-order moments of the observed signals, and are only constrained in that they cannot have poles or zeros on the unit circle. Krongold and Jones [97] developed similar rules that simultaneously minimise multiple snapshots of the cross-correlations while maintaining a constant-modulus constraint.

3.5.1.4 Maximum Likelihood Estimation

MLE approaches have also been applied to the separation of convolved non-stationary signal by Koutras *et al.* [98, 99], by considering short windowed sections of the signals, in which their characteristics were assumed to be stationary. They report significant improvement in automatic speech recognition performed after the separation.

3.5.2 Frequency domain approaches

While many of the initial approaches to tackling this problem employed a straight extension of the instantaneous solutions to each of the delays or taps to be considered, these are computationally demanding and would not be practical in situations where the lengths of filters used to model the responses are large. Since these lengths are dependent on the sampling frequency being used, lengths of many hundreds or even thousands would not be out of the ordinary in the modelling of a standard room due to the typical echo times measured — around 35 ms or greater, which, at 8 kHz (telephone quality) corresponds to 280 taps and upwards, and at 44.1 kHz (CD quality) corresponds to 1544 taps and beyond.

In order to attempt to reduce the computational load of this processing, some of the more

recent approaches have transformed the processing to the frequency domain, where many of the calculations are simplified. There are now many approaches to the separation and deconvolution of convolutively mixed signals that are carried out either partly or wholly in the frequency domain. These again fall into a number of different categories, listed below.

3.5.2.1 Transformed time domain approaches

The convolutive mixing problem can be transformed into the frequency domain by means of a Fourier Transform, where it becomes a set of independent instantaneous problems, one for each of the frequency bins generated by the transform. These independent problems can now be solved using any of the techniques discussed previously in Section 3.4. However, since each of these independent solutions may result in an arbitrary permutation and scaling, these must be unified before the solution is inverse-transformed back into the time domain.

This approach is adopted by Mejuto *et al.* [100], who use a separating technique based on Bell & Sejnowski's Infomax algorithm [10], and an evaluation of the fourth order cross-cumulant of the outputs to solve the permutation problem. Ikeda and Murata [101] used a decorrelation algorithm to resolve the ICA problem, rescaled the outputs using the corresponding elements of the inverses of the mixing matrices and then compared the envelopes of the different signals in all frequency bins to overcome the permutation problem. They note that other discontinuity criteria have also been successfully used by other authors.

3.5.2.2 FIR Polynomial Matrix Algebra

Lambert's FIR Polynomial Matrix Algebra [91] greatly simplifies the extension of learning rules for the separation of instantaneous mixtures to those capable of dealing with convolved mixtures, by easing the transformation to and from the frequency domain. It facilitates the manipulation of arrays of FIR filter coefficients by defining a class of polynomial matrices (matrices of polynomials), and mappings between the two. A set of algebraic operations that can be carried out are provided, and the basic properties of these elements are defined. This approach is illustrated by Lambert [102], where several learning rules applicable to the blind separation and deconvolution field are developed, including finite difference and serial update methods based on Amari's work. Lee *et al.* have also used it extensively in [59, 90, 92] and other related papers.

3.5.2.3 Frequency domain information-maximisation

Smaragdis [103, 104] has investigated the reformulation of infomax-based separation rules into the frequency domain, using Lambert's FIR Polynomial Matrix Algebra. It was reported that the transformation to the frequency domain permits a more efficient implementation of the algorithms, and eliminates many of the local minima problems due to temporal correlations. However, this transformation does not eliminate the ambiguity problems of scaling and permutation. The scaling problem can be overcome by normalising the mixing matrices found for each frequency bin, but no robust solution to the permutation problem has yet been found. This permutation problem leads to poor performance of the solution on more complicated filters using exponentially decaying Cauchy noise to simulate room impulse responses.

3.5.2.4 Frequency domain decorrelation

Parra & Spence [105] explain that simultaneous diagonalization of cross-power spectra of non-stationary signals at multiple time delays provides sufficient information to allow their separation from instantaneous mixtures. They also identify the difficulties in estimating the necessary cross-power spectra at the desired resolution when the stationarity time of the signals is short. A solution to the permutation problem can be achieved by imposing a constraint on the length of the filters with regard to the desired resolution of the DFT — that the filter size be comparatively shorter than the frame size — and requiring that the later taps of the filter are all zero. A fast-converging, on-line, frequency domain algorithm is described in [106] that makes use of some of these features, and for real-room recordings, with filter lengths of 2048 taps, converges within 3–6 seconds.

In [107] Fancourt & Parra describe the use of *coherence functions*. These coherence functions are a measure of the signal decorrelation between all possible pairs of outputs. An update rule, based on these measures, is derived to allow the separation of convolved sources — however, since this rule only separates, but does not deconvolve, the sources can only be recovered up to arbitrary permutation and filtering. This method is of low complexity and fast convergence, and is reported to offer good performance on speech signals.

3.5.2.5 CoBliSS

The Convolutive Blind Signal Separation (CoBliSS) algorithm by Schobben & Sommen [108] attempts to minimise the cross-correlations between the outputs in the frequency domain. The filters are then transformed back into the time domain, where due to the constraint of requiring a realisable filter, the new cross-correlations are no longer zero. Hence the algorithm iterates between satisfying these two constraints in the different domains. This algorithm makes use only of second order statistics, and requires no parameter tuning. The authors report fast convergence and good separation performance. In [109] the algorithm was extended to facilitate its use in a teleconferencing situation, and was tested on real signals with good results.

3.5.3 Summary of approaches

The key aspects of the approaches described have been summarised in Table 3.3 for ease of reference.

	Method	**Key aspects**
Time Domain	Information Maximisation	Builds directly from popular instantaneous solution, but poor performance on temporally correlated signals
	Natural Gradient with Non-holonomic Learning Constraints	Combination of two powerful but complex techniques
	Decorrelation	Can exploit multiple time delays to avoid estimation of higher order statistics
	Maximum Likelihood	Extension of the instantaneous case
Frequency Domain	Transformed Time Domain approaches	Reduce complexity by making a set of independent problems
	FIR Polynomial Matrix Algebra	Simplifies expression of learning rules and manipulation of data
	Information Maximisation	Poor performance due to multiple permutation problems
	Decorrelation	Fast convergence and low complexity
	CoBliSS	Fast convergence, no parameter tuning required

Table 3.3: *Summary of key aspects of Approaches to Blind Deconvolution*

3.6 Aims

The research in this thesis focuses on the linear, instantaneous and convolutive, stationary mixing of non-stationary, leptokurtic (speech) signals using relatively simple, general algorithms. It aims to demonstrate whether or not low computational-cost modifications can significantly improve the performance of simple, general solutions, and thereby offer potential benefit to practical implementations of speech separation systems.

The choice of speech as the type of signal for experimentation was made due to the current interest in speech-based applications. Linear, stationary mixing environments were selected to facilitate the identification of experimental effects in the results and minimise the number of other factors to be considered. The algorithms and techniques used were selected to keep the computational complexity of the system low, thereby making the contribution of the modifications more noticeable.

3.7 Summary

This chapter describes the blind separation and deconvolution problems, applications of the solutions to these problems and the various issues that must be addressed by these solutions. A variety of solutions to both the instantaneous and convolutive mixing problems are presented and different categories of the solutions are identified.

Chapter 4

The Effect of Signal Non-Stationarity on the Performance of Information-Maximisation-Based Blind Separation of Speech Signals

This chapter outlines the rationale for this part of the research, reviewing the existing work in this area, and looking at the reasons for the detrimental effect of signal non-stationarity on the separation performance of information-maximisation-based techniques. It defines a method for assessing the degree of non-stationarity of a signal, describes different methods of non-stationarity reduction, and describes experimentation undertaken that investigates the effects of one of these — silence removal — on the time domain based separation performance of instantaneously mixed signals with varying degrees of non-stationarity. Finally, conclusions are drawn and potential areas for future work are outlined.

4.1 Introduction to the research

Bell and Sejnowski's information-maximisation-based solution [10] to the blind signal processing problem provides an effective information theoretic framework upon which to base blind separation solutions. The algorithm performs very well when separating non-Gaussian, stationary sources in situations where there are a known number of such signals. In truly blind situations, however, where the number, type and extent of the input signals are not known, its performance can degrade substantially. In particular, it has been noted that the algorithm does not perform well at separating mixtures of non-stationary, non-Gaussian signals [53].

Existing research on blind separation of non-stationary signals has already been summarised in Section 3.2.1.3, and the main approaches considered in this work are noted again for ease of reference :

Time-dependent learning rules These actively use the signal non-stationarities to drive the network weights in the direction of a stable solution.

Nonholonomic orthogonal learning constraints This approach prevents updates that would lead to unstable solutions by limiting the directions in which updates can be be generated.

Pre-filtering The application of a simple low-pass filter to the inputs can be used to reduce the effect of rapid transients in the signal amplitude.

Intermittent adaptation Updates to one or more of the weights are carried out conditionally, based on an assessment of the output signals.

This chapter investigates a technique that, for the purposes of this study, is applied to the source signals prior to their mixing, to allow the effect of the technique to be assessed. This is developed in Chapter 5 to incorporate it into the separation system where it assesses the output signals and controls the update of the network, based on this assessment, in a manner similar to that of the intermittent adaptation.

Even though the number of works that address issues related to non-stationary signals has increased recently, there are still fewer publications in the area of blind separation than there are dealing with many of the other facets of the blind separation / ICA problem. This is particularly surprising considering that speech signals are often used as the sources to be separated, without reference to these characteristics (inherently non-stationary — and non-Gaussian — when the data comprises periods of time greater than 20–25 ms [110, 111]). The paucity of techniques that consider these classes of signals in the blind separation problem has, to some extent, limited the improvements achievable by developments in application areas such as telecommunications, particularly for tasks such as noise reduction. Now that there are a number of reliable frameworks for solving blind signal separation problems, attention has turned to practical application of the various techniques, to prove their worth in real-world situations.

Speech was selected as the source of experimental data as it is the focus of many communications applications, and is relatively easy to acquire. The effects of intrinsic speech signal characteristics on the performance of popular methods are examined in this thesis.

4.2 Reasons for performance loss

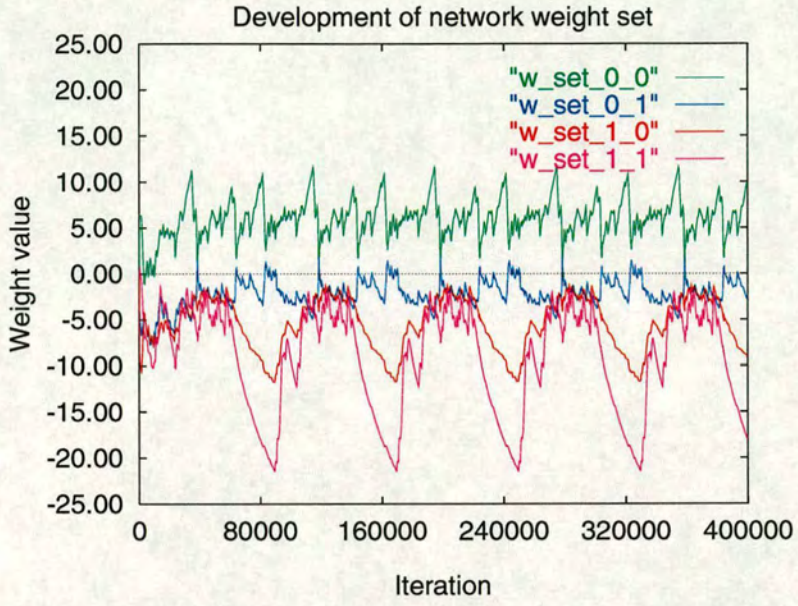
As can be seen in Section 3.4.3.2, Bell & Sejnowski's information-maximisation-based separation algorithm belongs to a class of neural network techniques governed by "gradient-based" methods. The performance of such techniques depends, to an extent, on certain characteristics of the network's inputs. In the case of blind separation, these factors include the degree of non-stationarity of the source signals, and of the mixing process. The changing characteristics of these features can alter the cost surface in the weight space dynamically, impeding convergence of the weight set to the desired solution. Experiments were performed to demonstrate this by comparing the development of the network weight sets, as shown in Figure 4.1. In the first case, Figure 4.1(a), the inputs were generated directly from speech signals that contain periods of silence, and hence have distinctly non-stationary characteristics at these points, particularly their variances. In the second, Figure 4.1(b), the inputs were generated from permuted versions of these sources and are therefore considered stationary. In these experiments, the mixing process was fixed and hence its stationarity was assured. The periodicity visible in the first of the graphs is due to cycling of the source data in the creation of the input signals.

It is the issue of the source signal non-stationarity, and its detrimental effect on the convergence of the network weight set that is addressed here in this thesis.

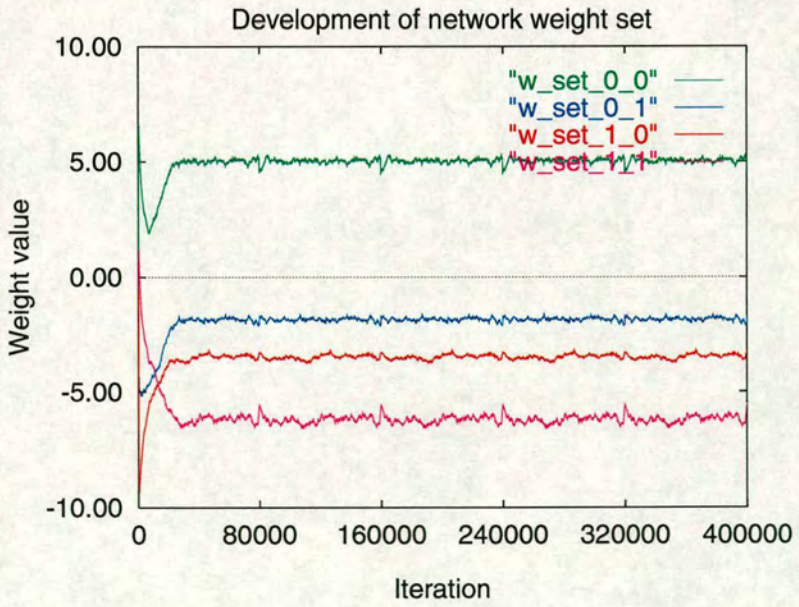
4.2.1 Susceptibility of the information-maximisation learning algorithm to signal non-stationarity

The reason that the information-maximisation-based blind source separation algorithm suffers in this way is that the changing characteristics of the signal cause a mismatch between the short term probability density function (PDF) of the signal, and the shape of the parameterised sigmoid function. Since the weight updates attempt to match this shape to the Cumulative Density Function (CDF) of the input signals, any such mismatches can cause the weight updates to diverge from the solution, thus leading to poor separation performance. If these changed characteristics last for more than a few updates, the system may start to converge towards this new target, from which it must then recover when the signal reverts to its original characteristics.

The most non-stationary characteristic of speech signals is their short term variance. This



(a) Original sources



(b) Permuted sources

Figure 4.1: *The effect of signal non-stationarity on weight set development*

fluctuates rapidly due to the bursty nature of speech. From an information-maximisation perspective, this can have a significant effect on the entropy of the outputs, which can be seen by consideration of their marginal entropy, $\widetilde{h}(Y_i)$. As noted in Haykin [18], the higher order cumulants of each of the outputs y_i, usually limited to κ_3, κ_4 and κ_6, must be estimated in order to evaluate this marginal entropy $\widetilde{h}(Y_i)$, which is used in determining the Kullback-Leibler divergence (see Section 3.4.1.2). These cumulants are defined in terms of the moments of the output signals, as estimates of the unknown source signals, according to the following relations :

$$\kappa_{i,3} = m_{i,3} \tag{4.1}$$

$$\kappa_{i,4} = m_{i,4} - 3m_{i,2}^2 \tag{4.2}$$

$$\kappa_{i,6} = m_{i,6} - 10m_{i,3}^2 - 15m_{i,2}m_{i,4} + 30m_{i,2}^3 \tag{4.3}$$

where $m_{a,b}$ is the b^{th} order moment of output y_a. It can be seen that rapid changes in the variance $\sigma_i^2 = m_{i,2}$ can result in marked changes in the higher order moments. The effects of these changes propagate through into the update rule for the network, and are what leads to the instability described by Amari *et al.* [54].

Haykin [18] also points out that the variance is unlikely to be constant, and highlights the choices for dealing with this observation :

- A unit-variance constraint can be enforced, which leads to inaccurate estimates for the higher order cumulants, and hence poor separation
- The variation in the variance can be viewed as a scaling of the whole signal, and the relative values of the moments and cumulants maintained. This so-called *unconstrained* approach leads to better separation results.

Due to the lack of quantitative data of the effect of this variation, it was decided to investigate the degree to which such non-stationarity affects performance, and possible means of overcoming this.

4.3 Signal non-stationarity

Before assessing the effect of non-stationarity on performance, first it is necessary to define "non-stationarity" for the context of this analysis. Definitions of the signal characteristics used

in this thesis can be found in Weisstein [55], so technical details are omitted here. However, it should be noted that whilst non-stationarity can take many forms, this research has focused mainly on the aspect most characteristic of speech signals — that of a relatively large short-term variation in the variance — since this is of primary significance in the convergence of the network weights to a stable separating solution. Measurement of the *degree* of non-stationarity of signals will also be defined in terms of these statistics. The techniques developed in this research will not be applicable to all other classes of non-stationary signals, since the signals will not necessarily exhibit these particular characteristics.

4.3.1 Measuring the degree of non-stationarity

To assess the degree of non-stationarity of the signals under consideration, it was necessary to define a *measure* of non-stationarity, since the term itself is normally used only to classify signals according to whether or not the various moments of the signal under consideration are constant throughout the duration of the signal.

Since the root of the convergence problems is the rapid variation in the amplitude of the signals and hence also in the running variance, it was decided to define the degree of non-stationarity of a signal with regard to this, at least in the context of these experiments. The method selected was based on the calculation of windowed statistics for the signal being assessed. An important point considered in the selection of this approach was the choice of an appropriate window duration over which to generate the statistics. The window must be short enough to follow any significant rapid variations, but long enough to allow the statistics generated to be meaningful. The window duration should be defined in terms of a period of time rather than a number of samples, to allow different sampling frequencies to be accommodated.

The resulting sets of mean and variance values were then assessed, themselves, to determine the range of their means and variances — after compensation for the dependency of the variance on the mean, which also kept the metric dimensionless, this gave a measure of how variable their values were over the length of the original signals, *i.e.* how non-stationary they were.

A formula that encapsulated the appropriate characteristics was derived :

$$\text{degree of non-stationarity} = \frac{\text{Var}\left(\text{Var}(x)\right)}{\text{E}[x]^2} \tag{4.4}$$

where E[x] and Var(x) denote the mean and variance of a set of values, respectively. The inner variance and mean terms are the set of windowed statistics, generated using the specified window period over the duration of the signal. The outer variance and mean terms show the characteristics of the set of the inner values, thereby allowing quantification of the variability of these characteristics. The measure tends towards zero for what are here deemed to be more stationary signals, as the variance of any moment of a stationary signal will be zero.

The *co-efficient of variation* (CV) is a measure of the spread of a set of data, relative to its mean. This makes it independent of the units of measurement. It is calculated as :

$$CV = \frac{\sigma}{\mu} \tag{4.5}$$

It can be seen that the measure of the degree of non-stationarity defined in Equation 4.4 is equivalent to the product of two co-efficients of variation, the inner one being time-windowed over the period of interest. This, then, is a measure of the variation in the signals' characteristics — a definition consistent with the desired quantification.

4.3.2 Methods of non-stationarity reduction

There are a variety of methods for reducing the non-stationarity of signals. Some take the form of mathematical manipulations such as whitening, others are based around basic filtering techniques, and yet more involve additional computational processing, such as permutation. A brief overview of some approaches is given here :

Permutation Permutation reduces non-stationarity in signals by scattering adjacent samples, thereby eliminating any time-dependent features of the data. The thoroughness of the operation depends on the size of the window over which the permutation is carried out, and on the method used to generate the random sequence of numbers for re-ordering the signal samples.

Bell and Sejnowski [10] used permutation to pre-process their source signals prior to mixing in their experiments, for exactly this purpose. Without such permutation, their separation results were not as good — a difference of the order of 10–15% over 50 seconds worth of 8 kHz speech data.

Iterative Compensation Iterative compensation involves maintaining estimates of the

moments of interest and using these to process the input signal at each step to produce an output that conforms to the desired constraint. An example of this is the subtraction of an estimate of the current mean from each sample of the incoming data to maintain a zero-mean signal.

Yang and Amari [112] have shown that this approach, or a re-scaling of a signal to maintain unit variance, can be incorporated into any linear mixture model, without loss of generality — hence this can be used in the blind signal separation algorithms dealt with in this research.

Filtering The objective of filtering, here, is to eliminate rapid changes in any of the lower order moments, most notably the mean and the variance. A low-pass filter can be used to achieve this, and was shown by Barros and Ohnishi [53] to have a beneficial effect on separation performance.

Silence removal Silence removal on speech signals can be used to eliminate the gaps between words, or even between phonemes within words, depending on the parameter values set for identifying the periods of silence. This reduces the severity of the variation in variance, thereby reducing the degree of its non-stationarity of the signal being processed.

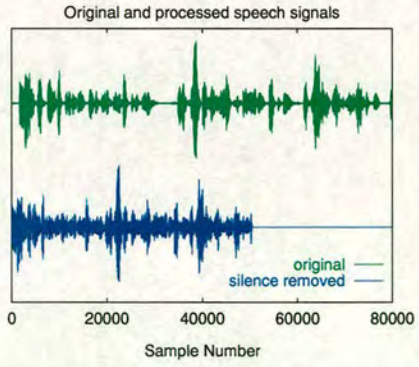
A number of methods may be used for the identification of silence within the signal. These are discussed more fully in Section 4.4.

Whilst other techniques exist that are capable of reducing the degree of non-stationarity exhibited by signals, many of them are far too complex to justify incorporation into a blind signal separation system. This is particularly true when computationally inexpensive algorithms are being sought, for reasons of modularity and implementation (see Section 3.1.4). For this reason, and the fact that the greatest divergence of the weights from the desired solution occurred during periods of silence in the source signals (see Figure 4.1(a)), the silence removal technique outlined above was adopted in this study as the method of choice for non-stationarity reduction. Silence removal, used frequently in digital telephony networks to reduce the amount of data that must be transmitted [113, 114], is also a practical problem itself.

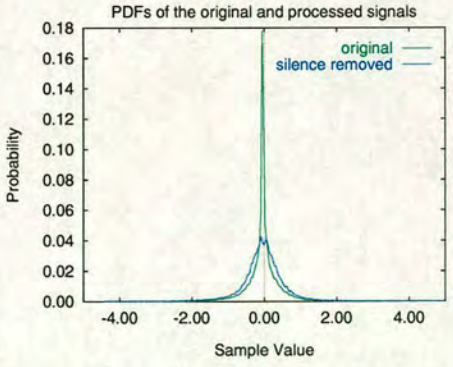
4.4 Silence removal

Although not generally applicable to all non-stationary signals, silence removal is of particular use with speech data since these signals tend to contain bursts of spikes, clustered together during the words of the speech, and separated from other words by gaps of relatively low signal amplitude. Hence, it is often relatively easy to identify the voiced sections from the original signals by inspection alone. Silence removal is used in a variety of speech-based applications, particularly to improve bandwidth usage in transoceanic telephone cables [113], and Voice Telephony Over ATM (VTOA) [114, 115], but neither of these make use of its stationarity-improving side-effects. The need to explore the applicability, beneficial or otherwise, of silence removal as a means of reducing the non-stationarity of speech signals is one of the principal reasons for this study.

A typical speech signal, such as that illustrated by the top graph in Figure 4.2(a), can be seen to be roughly zero-mean, with considerable variation in its amplitude, and consequently variance. (Modelling of this type of non-stationary process has been considered in [116].) The variation is greatest at the boundaries between regions of active speech and the gaps between them. Removing these periods of inactivity, henceforth referred to as "silence" (although this may not be strictly accurate), reduces the change in amplitude at that point, and the change in short-window mean value of the variance. This leads to a reduced degree of non-stationarity, in the context of this thesis. The results of such processing can be seen as the bottom graph of Figure 4.2(a). Figure 4.2(b) shows that the PDF is considerably less peaked after the processing.



(a) The original and processed speech signals



(b) PDFs of the original and processed signals

Figure 4.2: *The effect of silence removal on signal amplitude*

This is because the "silent" parts of the signal that are removed are grouped around its (near zero-) mean, making it appear more Gaussian. Bell & Sejnowski [10] illustrated the importance of PDF matching in their paper, and the appropriate selection of a non-linear function. While Haykin [18] comments that use of the logistic function limits the sources for which the rule can be effective to being super-Gaussian, this disagrees with Bell & Sejnowski's paper. In either case, the original speech signals used as the sources in this study are super-Gaussian.

In addition, the silence removal procedure can be extended to process the bursty parts of the speech, by removing the near-zero values between the spikes within the active speech. However, whilst this further reduces the variation in the signal variance, it also affects the characteristics of the active speech itself, leading to changes in its spectral characteristics and qualities such as pitch.

The critical parts of the silence removal process are the method by which periods of silence are identified, and the parameters by which these are defined — duration and threshold level. Selection of appropriate values for these parameters is discussed in more detail later.

4.4.1 Methods of silence identification and removal

There are a number of techniques that can be used to identify periods of silence within the signals. Some techniques are quite complicated and involve signal analysis such as counting zero-crossings [111], although more simple methods also exist. Most methods require specification of a period of time over which to assess the signal, and a threshold level, below which the signal is considered to be silent. The threshold value may be defined as an absolute data value, or as a term derived from this. Furthermore, the term need not be an instantaneous measure, and may take account of previous values of the signal.

Since simplicity of design was a consideration of this investigation, complex algorithms were not justified, or necessary. In this study, two relatively simple assessment schemes were devised and tested. One involved strict threshold comparisons between the signal data and the process parameters, and the second used an energy estimation technique intended to give the algorithm a better noise tolerance.

Method 1 : Strict threshold

The strict threshold comparison technique for silence identification involved moving a window of the desired duration (*i.e.* a number of samples) through the signal being assessed. Data was written (from the beginning of the window) to the output file only when the absolute value of one or more of the samples currently in the window exceeded the threshold level (*i.e.* when the signal was not silent). Selection of an appropriate duration is discussed in Section 4.4.2 and thresholding considerations in Section 4.4.3.

Whilst simple, this technique was reasonably effective, although experiments showed that it performed poorly when the threshold was set too low. This is because the comparison of the data is carried out without regard for the presence of noise at the desired threshold level. A single sample that breaks the threshold, regardless of by how much, causes the process to consider all windows containing that sample to be non-silent. This is illustrated below in Figure 4.3. Consequently, an alternative approach that incorporated a degree of noise tolerance

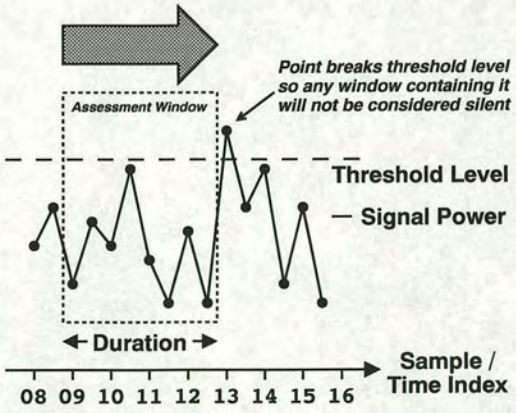


Figure 4.3: *Silence identification by strict threshold comparison*

was implemented and tested.

Method 2 : Average energy threshold

The second technique is referred to as the "average energy" method as it makes use of the mean value of the instantaneous power of the signal. Since a finite-sized window is being assessed,

the concept of energy is appropriate here [12] :

$$E = \int_{-\infty}^{+\infty} [x(t)]^2 dt \tag{4.6}$$

where the limits of the integration become the extents of the window size used. The resulting value is then divided by the window size to give an average energy value per sample. While this average value was below the specified threshold, no output was written.

The average energy technique is more tolerant of signal noise than the strict threshold technique, due to the averaging process. Small numbers of slight breaches of the threshold may now be tolerated without causing the entire windowed section to be classified as non-silent. However, significant intrusions that alter the mean value of the window's energy sufficiently to cause it to cross the threshold again lead to that window being classified as non-silent (see Figure 4.4). Such large changes are more likely to be the result of a change in signal state (*i.e.* a transition from a gap to a word in the speech signals), rather than noise.

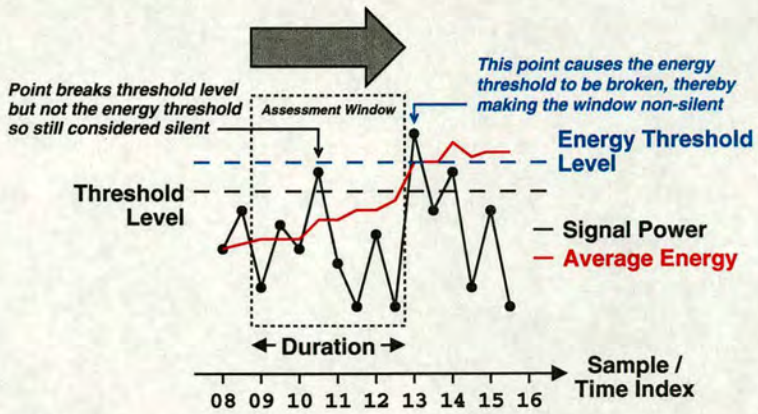


Figure 4.4: *Noise tolerance in the silence detection process by means of average energy estimation*

4.4.2 Duration of silence

Both of these silence assessment methods require the establishment of the optimal duration of the period of silence to be removed. The value chosen determines the window size over which the silence assessment is to be carried out. For the strict thresholding case, this period should be more accurately described as the optimal *minimum* period, as periods of silence longer than the specified duration will also satisfy the silence criterion. Once a suitable value is found, below

which there is no worthwhile gain in performance, longer durations need not be considered. It should be noted that the length of this period may vary with the method of identification used, and the threshold level set.

When using the average energy method for silence identification, the relevance of the parameter changes slightly. It is still desirable to remove only as little of the signal as is necessary, so that there is more data for the network to train on, but the period of assessment is now fixed. However, since the window of assessment is moved through the signal one sample at a time, the average energy method should, at worst, remove the same long periods of silence as the strict comparison method, plus extensions of these and potentially some other sections that do not satisfy the strict criteria. Thus, for a given duration (and assuming a set threshold level), the average energy silence identification method should never produce a longer output than the strict thresholding method.

The optimal duration determined for the average energy method may not be the same as the *minimum* determined for the strict thresholding method due to the non-uniform distribution of signal energy within the window at any instant in time, and the fixed versus expandable periods of assessment.

When considering the range of durations over which to perform silence removal, several factors relating to speech signals were considered :

Short window stationarity Speech signals are generally considered stationary over periods of 20–25 ms or less [110,111]. Within these periods, any apparent gaps or periods of silence are actually integral parts of the speech signal, and cannot be removed without affecting the signal's inherent characteristics.

Inter-word gaps Typical inter-word gaps can range from 50 ms to 800 ms, with the most frequent values in the range 250–500 ms [113]. The intention is to remove these inter-word gaps without affecting the key parts of the signal.

Sample rate The sample rate of the data must be high enough to ensure that no aliasing of the signal has occurred during the recording [117]. Other than this, the main considerations that need be paid to the sample rate are that the window duration and the signal sizes are dealt with in terms of numbers of samples.

To a certain extent, the optimal durations determined will be dependent on the actual speech

signals used. This can be attributed to the speaker's style, and whether they are reading aloud or talking freely. However, the studies in this thesis provide general guidelines on the order of magnitude of window size that lead to noteworthy changes in separating performance.

4.4.3 Threshold of silence

The final question needing consideration was that of determining a suitable threshold level at which to classify the signals as silent or not. This level need not be defined directly in terms of signal data values, but may instead be based on some function of them. Since the signals to be dealt with may vary considerably in range or scale, setting a fixed absolute threshold would not be appropriate, as it could lead to mismatched identification of periods of silence between the different inputs.

Alternative solutions exist, such as the definition of a relative threshold. This particular method, which must adapt to the input signal, scaling itself with the maximum range of the signal observed so far, was selected for use in the subsequent experiments. Ideally, some scale-independent threshold measure should be used, such as an entropy-based criterion. Such measures are impossible to calculate accurately, since the entire signal is not known in advance, but an approximation of the relative entropy over the window size of assessment could be calculated instead. This value could then be compared to the maximum or average value, or even to some pre-determined fixed value derived from typical statistics for speech signals.

To facilitate comparison between the results of the two methods of identification described earlier, the threshold levels for both techniques were defined in terms of a specified percentage of the maximum possible instantaneous signal energy. The formula used to calculate the threshold was :

$$threshold = \left(\frac{l}{100} * max\right)^2 \tag{4.7}$$

where l is the level below which the signal is to be deemed silent, as a percentage of the absolute maximum signal value max. The value is squared to give the instantaneous power, according to :

$$P_{inst}(x(t)) = x(t)^2 \tag{4.8}$$

It should again be noted that in the case of the average energy method, this level is the threshold against which the average energy of the whole buffer is compared, rather than that of the individual samples, as is the case when using the strict thresholding method.

4.5 Separation assessment criteria

To assess the effect of the varying degrees of non-stationarity of the source signals on separation performance, a method of quantifying the degree of separation attained was required. This assessment need not form part of the experimental simulation, and can be performed independently, allowing various techniques to be tried without having to re-run the experiments.

One way of assessing the quality of the separation achieved is by listening to the output signals. However, this is very subjective and not at all quantifiable. It can still give a useful indication of the progress of the separation and is arguably more useful than a range of figures in determining when an 'acceptable' degree of separation or improvement has been achieved. This assessment is further aided in the controlled experimental setup by the availability of both the original and the processed source signals, which can also be listened to.

In the experimental work, as full knowledge of the mixing matrices used was available, recording the values of the weight set during the experiments allowed an analysis of how effectively the network's weight set had converged. This is a useful assessment, since the values of the matrices involved have clearly definable relationships and properties. For example, the product of the original mixing matrix and that representing the network's weight set should yield a scaled, permuted identity matrix when a desired solution has been achieved, as explained in Section 3.1.1.1, Equation 3.3.

Several suggestions for generalised performance assessment metrics have been proposed in the past, including those noted in [53, 58, 118]. Various metrics were considered for use and two, that of Amari *et al.* [58] and that of Barros and Ohnishi [53], were selected due to the differences in the properties they quantify.

The descriptions of these two separation performance assessment algorithms will use the following labels :

A is the mixing matrix

W is the network's weight set, that approximates the desired separating (or 'unmixing') matrix

C is the product of these two matrices, a scaled, permuted identity matrix

Elements of these matrices are indexed in the usual format, *e.g.* c_{ij} is the j^{th} element of the i^{th} row of the matrix **C**.

4.5.1 Amari *et al.*'s performance metric

The first of these performance metrics, set out by Amari *et al.* in [58], uses the fact that the product matrix **C**, despite potentially being scaled and permuted, should still have only one non-zero entry per row and column. The scaling in the matrix will be independent by row or column, and so can be accounted for by normalising the values in each row or column with respect to their sum. The permutation need not be resolved since the sum of the values in any row or column, once they have been appropriately normalised, should still be one. Consequently, subtracting one from the sum indicates the degree to which that row or column has achieved its solution. Summing over all rows or columns therefore gives an indication of how close the matrix **W** currently is to a full solution . This is encapsulated in [58] by the formula :

$$E = \sum_{i=1}^{n} \left(\sum_{j=1}^{n} \frac{|c_{ij}|}{\max_k |c_{ik}|} - 1 \right) + \sum_{j=1}^{n} \left(\sum_{i=1}^{n} \frac{|c_{ij}|}{\max_k |c_{kj}|} - 1 \right) \tag{4.9}$$

where i, j and k are indices that all range over the size of the square matrix **C**, and E is then a measure of the overall residual error in the system.

4.5.2 Barros & Ohnishi's performance metric

The second method considered is that proposed by Barros & Ohnishi in [53]. Based on a similar formula to Amari *et al.*'s metric above, Barros & Ohnishi's metric defines a performance measure for each output, describing how close the maximum normalised value in a column is

to unity. They propose the formula :

$$p_j = \max\left\{ \frac{|c_{ij}|}{\sum_j |c_{ij}|} \right\} \Bigg|_i \quad \forall i, j \tag{4.10}$$

where p_j is the performance of the j^{th} output, and the other variables are as described previously. The values of these performance metrics will always be in the range $0.5 \leq p_j \leq 1.0$, where 0.5 indicates no separation and 1.0 indicates full separation. (As presented above, the formula could lead to misleading output traces if the product matrix diverges too far from an idealised solution and has the maximum values of each row in the same column, or vice versa. However, this can be overcome by ensuring that once a value has been picked, its row and column are eliminated from the remaining set.)

Having a measure of performance for each output is helpful in identifying interesting regions of the input signal, that have lead to sudden changes in the rate of convergence towards the solution. This particular method also allows clear indication of how the separation of each individual signal is progressing.

Both Amari's and Barros & Ohnishi's performance metrics are used in subsequent assessments in Section 4.7.

4.6 The effect of signal non-stationarity on the performance of information-maximisation-based blind separation

Having set out the theory behind the techniques used in this investigation, the experiments carried out are now described, results presented and their implications discussed.

4.6.1 Blind separation experiments

To demonstrate the effect of signal non-stationarity on the separation performance of the information-maximisation-based blind separation algorithm a simple comparative experiment was undertaken. Two simulations were carried out using identical network configurations and setups, one with inputs created by direct mixing of the original source signals under consideration — in this case two BT speech signals, here labelled **BT_female** and **BT_male** — and the other with inputs created from mixtures of time-permuted versions of these same

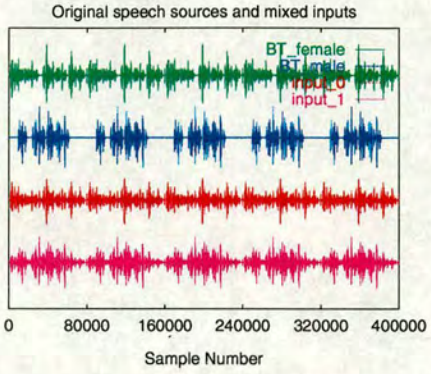
sources. In the latter case the order of the samples within each of the source signals was permuted independently, and then the inputs created by mixing corresponding samples from these re-ordered sources in the same way as before, as initially described by Bell & Sejnowski [10]. This permutation has the effect of eliminating any non-stationary characteristics of the signals.

Both simulations were started from the same state (*i.e.* initial weight and bias values) and were run for 400 000 iterations, equivalent to five passes through the input data. Sections of the source and input signals are shown in Figure 4.5, along with the development of the weight set values over the course of the experiment and the corresponding performance assessment using Amari *et al.*'s metric [58]. The periodicity visible in the graphs is due to cycling of the source data during the creation of the input signals. The weights of the first network, processing the inputs from the original sources (Figure 4.5(c)), show large changes in value in each of the weights, throughout the duration of the simulation. These large changes in weight value result in a lack of convergence to stable values. The separation performance graph (Figure 4.5(e)) also shows that these variations have prevented convergence to a stable solution.

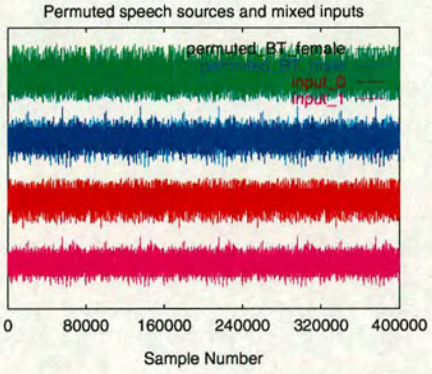
The weights of the second network, processing the inputs generated from the time-permuted sources (Figure 4.5(d)), converge smoothly and rapidly to stable values. The graph of the value of Amari's performance metric (Figure 4.5(f)), calculated at each batch update of the weight set, shows that the state reached is a good separating solution of this problem.

Looking at the graphs in more detail, it can be seen that the large variation in value of weight w_{11} in Figure 4.5(c) clearly corresponds to the periods when **BT_male** is quiet, and weight w_{00} varies similarly with **BT_female**. These weights are on the leading diagonal of the matrix which, in these experiments, is where the separated signals also appear, and thus show the largest effect of these changes. However, the off-diagonal weights w_{01} and w_{10} also show similar variations which again can be traced to the periods of quiet in the source signals. It was this observation that lead to the decision to follow this line of investigation.

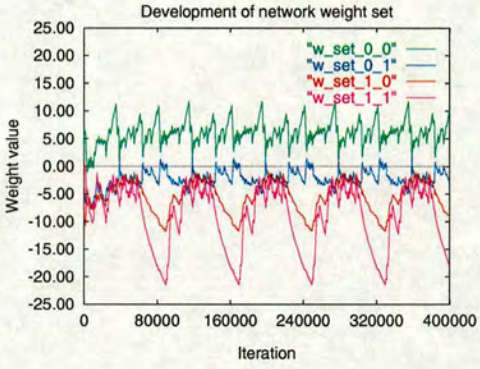
An interesting point to note is that it is difficult to identify the periods of the *input* signals that correspond to the quiet parts of the sources, so that in a truly blind system no direct pre-processing of these inputs could be performed to eliminate potential problem sections. For an on-line implementation of this processing, a method capable of identifying these sections of the unknown source signals would be needed so that the convergence of the network's weight



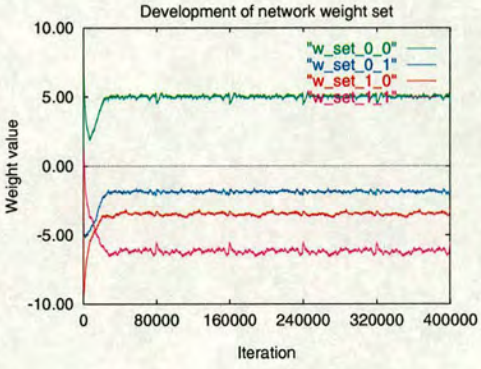
(a) The original source and input signals



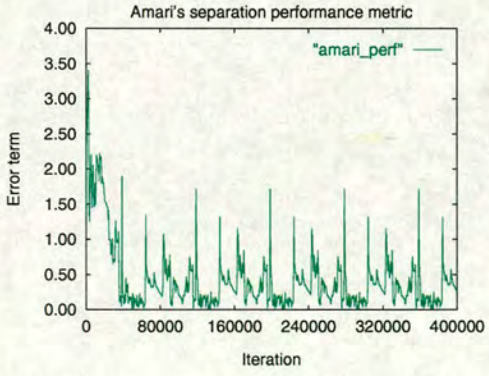
(b) The permuted source and input signals



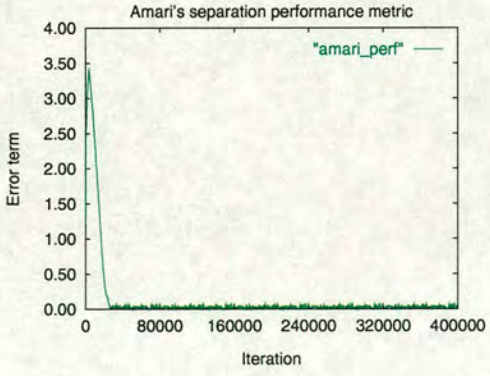
(c) Development of the network weights for original inputs



(d) Development of the network weights for permuted inputs



(e) Amari's separation performance metric for the original mixtures



(f) Amari's separation performance metric for the permuted mixtures

Figure 4.5: *Illustration of effect of non-stationarity on weight set development and separation performance*

set would not suffer. This problem is addressed in Chapter 5.

4.6.2 Investigation of the sensitivity of the separation algorithm

The aim of this part of the research was to assess the sensitivity of the information-maximisation blind separation algorithm to the non-stationary signal characteristics exhibited by speech signals. A range of experiments was designed to determine whether appropriately parameterised silence removal techniques could be used to control the degree of non-stationarity of the signals, as defined in Section 4.3.1. Results generated will illustrate the variation in performance of a blind signal separating system based on Bell & Sejnowski's information-maximisation-based algorithm, when run on mixtures of signals of various degrees of non-stationarity. The effect of batch-sizing was also investigated.

4.6.3 Experiment design

The first part of this investigation was carried out to determine whether the removal of periods of silence of varying durations and with a range of thresholds produced a controllable change in the degree of non-stationarity exhibited by speech signals.

The next set of experiments was designed to investigate the relationship between the degree of non-stationarity of source signals, linearly mixed to create input signals, and the separating performance of the information-maximisation-based blind separation algorithm. In addition to this, the relationship between the degree of non-stationarity and the silence removal parameters would be determined.

Finally, the effect of different batch sizes (see Section 2.2.3.1) on the separation performance was assessed, since the number of iterations over which the updates are summed correspond to time intervals within the signals, determined by the signal's sample rate. Consequently, this summation may be affected by the inherent non-stationarity characteristics of the signals, and may in turn affect performance.

4.6.4 Selection of source data

The focus of this part of the thesis is the effect of silence removal on the performance of algorithms for the blind separation of signals which exhibit non-stationary characteristics, and

in particular those typical of speech signals. The principal characteristics of speech signals that made them suitable as the source data for the experiments were :

Non-stationarity Speech signals exhibit non-stationary characteristics — most notably a rapidly changing variance, at least when considered for periods longer than about 20 ms.

Availability In general, speech signals are easy to obtain, simply by recording a test subject speaking and then digitising the recorded speech, or by direct digital recording. A number of options are available in this process, allowing control over aspects such as the sample rate and the resolution of the data.

It was desirable to have available a number of different source data signals, exhibiting a variety of different characteristics, to allow various combinations of these to be mixed as inputs. This meant that the silence removal techniques, and later the blind separation techniques, could be better assessed for generality.

BT Laboratories, Martlesham Heath, were approached to supply some of the data. The data provided by BT Labs had been sampled at 8 kHz from various DAT recordings, with 12-bit resolution and a bandwidth of 3.3 kHz (*i.e.* telephone quality speech), using a linear array of fifteen microphones. Initially, the intention was to use this data to create the mixed input signals, simply by adding together corresponding (optionally scaled) microphone outputs from the various data sets. However, to allow better control in the experiments that follow, only the output from the centre microphone was used. Therefore, when the signals were artificially mixed, using a known mixing matrix, the exact proportions of each source signal in the inputs generated would also be known. This gave more control over the experimental setup, and allowed better analysis of the separation test results. From the data provided, two signals were selected - a female voice reading a news article, and a male voice reporting an error. Both of these signals were exactly ten seconds long.

Other data sets for use in this thesis were generated in-house, by directly recording different test subjects in an anechoic chamber, to eliminate time-delayed echoes and hence keep the "blind" part of the problem limited purely to separation, rather than a combination of separation and deconvolution. The recordings were made on a digital recorder, so there was no need to subsequently digitise the samples, and were of a higher quality than the BT data sets, being sampled at 44.1 kHz, with 16-bit resolution. The subjects were asked to talk normally, as if in conversation, to ensure as natural sounding speech as possible. This should eliminate

any artificially introduced timing changes, such as may occur when reading aloud from a script. A range of recordings were collected. Ten second samples were selected from three of the recordings — one female and two different male voices — and converted down to an 8 kHz sample rate, after appropriate filtering. This was necessary to permit mixing between the different data sets.

All of the signals to be used in the experiments were then scaled so that their sampled data values all lay within the range $[-5.00 : 5.00]$. This procedure ensured that for mixtures of signals drawn from the two different data sets, the signals were better matched with one another in terms of amplitude. The different resolution would otherwise have resulted in signals from the second set swamping those from the BT data set, making separation particularly difficult. The exact scaling is not critical to the solution, and the source signals are presented for reference in Figure 4.6, to show their shape :

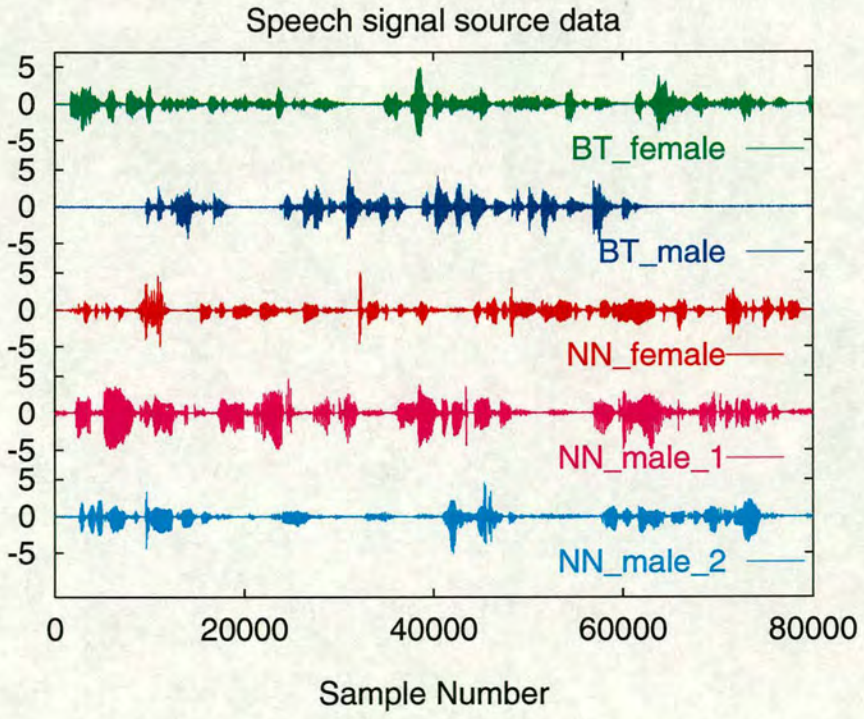


Figure 4.6: *The scaled source speech signals*

4.6.5 Creation of input data

The input signals for the simulations were created by artificially mixing data from the source signals described above. Unlike the data described in Bell & Sejnowski's paper [10], the sources were *not* permuted prior to this mixing, but were pre-processed using the two silence removal techniques described in Section 4.4.1 — the strict threshold technique and the average energy technique. This pre-processing was carried out with a range of both duration and threshold parameters, details of which can be found in Section 4.7.1. The mixing was carried out according to the basic mathematical formula given in Equation 3.1.

A range of mixing matrices (see Table 4.1) were used, of varying determinant magnitude, such that some were well-conditioned, and others were nearer singular and would therefore generate mixtures that were more difficult to separate. Since the mixing matrix used in each experiment was known, the separation performance could be related back to the conditioning of the mixing matrix, as well as the degree of non-stationarity of the source signals used to create the inputs.

The experiments were designed to be as uncomplicated as possible, and had only two inputs and two outputs. Therefore, the matrices used each comprised of two rows of two columns. These were generated from random numbers uniformly distributed in the range $[-1.0 : 1.0]$, with the exception of matrix **A6** which was designed to have the maximum possible determinant, of value 2.00. The matrices are presented in Table 4.1, and their determinants in Table 4.2.

$\mathbf{A1} = \begin{bmatrix} 0.19534 & -0.44666 \\ 0.32684 & -0.29977 \end{bmatrix}$	$\mathbf{A2} = \begin{bmatrix} -0.45789 & 0.49435 \\ -0.17221 & -0.41647 \end{bmatrix}$
$\mathbf{A3} = \begin{bmatrix} 0.80600 & -0.36289 \\ 0.65076 & -0.93591 \end{bmatrix}$	$\mathbf{A4} = \begin{bmatrix} 0.27685 & 0.76390 \\ -0.99081 & 0.13659 \end{bmatrix}$
$\mathbf{A5} = \begin{bmatrix} -0.93488 & -0.80780 \\ 0.47754 & -0.97064 \end{bmatrix}$	$\mathbf{A6} = \begin{bmatrix} 1.00000 & 1.00000 \\ -1.00000 & 1.00000 \end{bmatrix}$

Table 4.1: *The 2 by 2 mixing matrices*

The desired length (in iterations) of the input signals was specified in the mixing process, and the mixture-generating code looped back independently through each source signal as required,

$\det(\mathbf{A1}) =$	0.08743	$\det(\mathbf{A2}) =$	0.27583
$\det(\mathbf{A3}) =$	−0.51820	$\det(\mathbf{A4}) =$	0.79469
$\det(\mathbf{A5}) =$	1.29319	$\det(\mathbf{A6}) =$	2.00000

Table 4.2: *Determinants of the 2 by 2 mixing matrices*

until the necessary number of samples had been processed. This method ensured that if the pre-processed source signals were of different lengths, the start of the second pass through one of them would not coincide with the start of the second pass through the other.

Although preliminary tests indicated that most of the separation actually occurred within the first pass through the data, it was deemed appropriate to run the simulations for five times this length to allow the separation metric to settle to a steady state.

In practice, to allow for a random time offset into the input signals, for averaging purposes, the length of the mixed data generated was actually six times that of the sources. The original source data signals were all precisely ten seconds long, and were recorded at 8 kHz. This meant that the data signals each contained 80000 samples. Consequently, each of the mixed input signals generated from these sources was 480000 samples long — see Section 4.7.3.

4.6.6 Experimental setup

A parameterisable software model of a blind signal separation system, based on the single layer neural architecture described in Bell & Sejnowski's paper [10] was designed and used for the simulations. This new model was modified to allow variation of a number of the control parameters for each simulation, including :

Number of input and output signals Whilst the number of outputs always matched the number of inputs, both of these could be altered independently to allow testing of larger networks. In the initial experiments, both of these parameters were set to two.

Duration of the simulation The total running time of the simulation could be specified so that if a particular mixture was proving hard to separate, the simulation could be given longer to process it. However, in the present studies, this value was fixed at 400000 iterations for the reasons explained previously. (See Sections 4.6.5 and 4.7.3.)

Time offset A time offset into the inputs signal could be specified to allow results to be generated starting at a different point in the signals. The offset used was the same for all signals in a particular simulation run, so that corresponding samples were aligned in time. This facilitated averaging of the results over time, in an attempt to smooth out time correlated features of the signals that could obscure the true results.

Batch size The number of iterations over which the updates generated were summed prior to updating the weight set and bias vectors was also controllable. This meant that the relationship between signal characteristics and separation performance could be investigated.

Network initialisation The design enabled the initial values of the network's weight set and bias values to be set by selection from a range of configuration files. This allowed results from several different start points to be generated, to demonstrate that they were not dependent on particular starting conditions. The configuration files contained values generated randomly, from a uniform distribution of range $[-1.0 : 1.0]$.

Separation / Update algorithm With only minor modification, the network could be reconfigured to use alternative separation techniques, and update generation rules. This capability allowed comparison of the results between techniques. For the first set of experiments, only Bell & Sejnowski's infomax algorithm was used. In subsequent experiments, reported in Chapter 5, algorithms by Jutten & Hérault, Matsuoka, Ohya & Kawamoto, and Barros & Ohnishi were also used.

Output generation The design of the model facilitated the outputting any of the values of interest connected with the separation process. This meant that the outputs, weight set, bias values and various other internal values could all be monitored, as desired, during the progress of the simulation.

Code for the model was initially written in C++, but subsequently re-written in C as this enabled a considerable speedup by the elimination of the object-oriented overhead. Modular coding techniques were employed wherever possible in all code produced, and appropriate testing was carried out. A suite of shell scripts were developed, tested and employed to schedule large batches of simulations with a range of five threshold levels, five durations, six mixtures and thirty different starting conditions, and then to collate and process the results afterwards to generate the performance metrics previously discussed in Section 4.5.

The focus of the experiment was on the separation performance, and due to the separation assessment criteria discussed previously, the principal interest was the development of the network's weight set. For this reason, the output signals were used only for a subjective assessment of the audible quality of the separated signals, to confirm the performance metrics' evaluation of the separation. The weight set values were recorded after every update (*i.e.* once per batch rather than once per simulator iteration). Once each simulation run had finished, the weight sets were processed using Amari *et al.*'s metric [58], and in some cases the metric of Barros & Ohnishi [53] as well, and graphs generated of the separation performance. The initial graphs produced were averaged over the six different start times for a particular initial weight set configuration to allow any trends in the performance over the course of the simulation to be identified. It was recognised that this averaging over the time offsets may results in noisy graphs, but the use of the windowed statistic method described in Section 4.3.1 allowed a final value to be extracted from the mean generated, for use in direct comparisons. In subsequent runs undertaken for the statistical analysis, the output data was averaged over all of the initial weight set configurations and all of the start times. Although this would average out trends from particular initial weight set configurations, it would provide a fuller picture of the overall performance.

4.6.7 Statistical analysis

An Analysis of Variance (ANOVA) was carried out on each set of results, using Genstat [119]. The different mixtures, duration and threshold values and, where appropriate, buffer sizes, were all used as factors in the analyses. Thirty replicates were achieved by starting the simulations from the five different initial weight set configurations and the six time offsets. This meant that the results gave an overall picture of the performance from the range of starting conditions. The full ANOVA tables for the experiments considered can be found in Appendix B.

4.6.8 The effect of batch sizes

To reduce the processing time for each simulation, the update algorithm used was coded to make use of *batch processing* techniques, whereby the computationally expensive weight set updates to the network were calculated much less frequently. Bell & Sejnowski [10] briefly mentioned the selection of appropriate batch sizes and learning rates, in their paper, but did not provide any details. In their experiments they suggested a usable range from 5 to 300 updates

and opted for a value of 200, with a learning rate of 0.01.

Since the batch size used corresponds to a number of samples, and hence a period of time in the input signals, the characteristics of the signal over that period must be considered when deciding on batch size. 200 samples at 8 kHz, which is the rate at which Bell & Sejnowski's signals were sampled, resolves to a time period of 25 ms, very close to the accepted duration over which speech signals may be classed as non-stationary. A range of experiments was devised to determine the how batch size affected the separation performance. The range of batch sizes selected was chosen to cover this critical duration of 25 ms. The learning rate was kept constant at 0.01 so that only the effect of the changing batch size would affect the results.

The source signal characteristics, over the duration of the shorter batches, would be classed and considered as stationary. However, for the longer batch sizes, they will exhibit a degree of non-stationarity in their variance, and this will be incorporated into the values of the updates generated during the batch processing. This will affect the dynamics of the learning algorithm, as the batch update value obtained may no longer lead in the same direction as would have resulted from summation of stationary values. This effect may impinge on the convergence of the weight set to the desired solution, but only over many batches.

In some cases, this change in dynamics may be advantageous, allowing escape from any local extrema that still exist in the non-stationary weight space, while in other cases it will undoubtedly lead to yet more variation from the desired convergence trajectory.

4.7 Experimental results

This section describes results from the experiments described above, performed on silence removal, blind separation and batch sizing and discusses the results in the context of this investigation.

4.7.1 Silence removal

The two silence removal algorithms described earlier (see Section 4.4.1), using strict thresholding and the energy equivalent thresholding technique, were run on the two test source signals described, **BT_female** and **BT_male**, using a range of parameters shown below. A wide range of values were tried with both methods on both sources and the resulting signals

examined.

From the set of processed signals, a subset of the parameters that resulted in a reasonable range of outputs were selected. These were 5 ms, 50 ms, 0.1 s, 0.5 s and 1.0 s. The low end of this range is below the stationarity threshold, and the upper end is above the normal inter-word length. Silence thresholds of 0.10%, 0.50%, 1.00%, 5.00% and 10.00% were selected.

The resultant signal lengths showed the trends in size expected from the silence removal processes. For any fixed threshold level, there are fewer long periods than short ones that satisfy the silence criteria, and consequently there is more reduction in output data size for short durations. Equally, for any fixed duration, higher thresholds result in more of the signal being classified as silent, and consequently removed. This again results in shorter signal lengths.

4.7.2 Non-stationarity assessment

To allow corresponding trends in the separation performance to be identified, it was necessary for the source signals to be assessed for non-stationarity. The assessment was made as described earlier using Equation 4.4, as in Section 4.3.1. A window size of 10 ms was chosen for the running window statistics, used in generating the short window variance of the signal as this size was well below the duration at which speech signals may be considered stationary and minimized any potential error due to non-stationarity in the statistics generated.

The values produced by this assessment are dimensionless and provide a relative guide to the degree of non-stationarity of the signal. The more stationary the signal, the lower the values, since the range of the variance will not fluctuate as much. Consequently, the anticipated *trend* is of decreasing values with reduced non-stationarity.

Since the non-stationarity of any of the sources affects the separation performance, the corresponding non-stationarity assessment figures from the two signals used in these experiments were added. This gave an overall assessment level for each set of simulations. These total values are given below in Tables 4.3 and 4.4, for the strict and the average energy thresholding techniques respectively.

Two trends can be seen in the data from the strict thresholding experiments (Table 4.3), each corresponding to one parameter of the silence removal process. For a fixed threshold level, the more stationary signals (smaller values) are to be found at the shorter durations. Conversely, for

		Duration (s)				
		0.005	0.050	0.100	0.500	1.000
Threshold (%)	0.10	13.42838	13.43395	13.43395	13.43395	13.43395
	0.50	9.92081	10.43724	10.46560	12.22006	13.43395
	1.00	9.39046	9.84611	9.95899	10.71508	11.61939
	5.00	6.41183	7.78821	8.63423	10.56101	11.57232
	10.00	4.30635	5.57867	6.69740	10.42164	11.56466

Table 4.3: *Sum of non-stationarity assessment for both signals, using* **strict thresholding**

		Duration (s)				
		0.005	0.050	0.100	0.500	1.000
Threshold (%)	0.10	12.90903	13.13474	13.24242	13.35684	13.25896
	0.50	9.48846	9.89850	10.26098	11.56470	12.09032
	1.00	8.60458	9.35476	9.99440	11.44632	12.08887
	5.00	5.04936	6.18313	7.54976	10.25117	11.13790
	10.00	2.41574	3.10907	4.50377	7.12072	9.67621

Table 4.4: *Sum of non-stationarity assessment for both signals, using* **average energy thresholding**

a fixed duration the more stationary signals coincide with the higher thresholds. Both of these results concur with the expected results, and lead to the fact that the most stationary signals will be achieved by setting the threshold high, and the duration low. The trend of increasing non-stationarity runs from the bottom left of the table to the top right.

Similar trends are seen in the results of the tests using the energy equivalent thresholding approach (Table 4.4), although the values achieved across most of the table are much lower, indicating that the outputs generated are more stationary. This result is as anticipated, due to the better noise tolerance of this technique.

These trends follow those found in the lengths of the output signal data, after the silence removal process, which is in accordance with the mechanism by which the statistics and metrics are generated.

4.7.3 Blind signal separation

Having generated the source signals of varying degrees of non-stationarity, corresponding pairs of signals, *i.e.* those processed using the same silence removal algorithm and with the same parameters, were mixed using each of the six mixing matrices previously described in Table 4.1, Section 4.6.5. (Pairs of signals processed with different parameters could equally well have been mixed, since the initial assessment will be based on the degree of non-stationarity of the two signals, rather than on the silence removal parameters. However, by keeping them the same the effect of these parameters on separation could also be studied.) Each mixture was made 480000 samples long, to allow a random offset of up to 80000, and still ensure that there was sufficient data for 400000 iterations of the separation process.

Each mixture of signals was passed through the network thirty times — comprising five random initial network configurations and six random time offsets — for each set of silence removal parameters. The results were used to generate the performance metrics discussed in Section 4.5. These metrics were then averaged over the twenty-five runs to give a single final set of results for each mixing matrix, and each set of parameters, using the selected silence removal algorithm. The graphs presented later are those derived from Amari *et al.*'s metric, since the single graph gives an overall indication of the separation performance of the whole system, for each run. Separate graphs generated for each output signal by Barros & Ohnishi's test (not shown) exhibited similar trends.

The graphs generated often contained significant amounts of noise, for example Figures 4.8, 4.9 and 4.13. Representative values of the separation performance were obtained by running a windowed mean over the graphed metrics, and selecting the final value of each such trace. Due to the noisy nature of the graphs, the figures obtained could serve only as a guide to the degree of separation achieved.

For further analysis, the data was subjected to an Analysis of Variance (ANOVA) using Genstat [119]. The final values of the asymptotes of each individual trace were compared, and the possibility of interactions between the parameters investigated.

Relationship between non-stationarity and separation performance

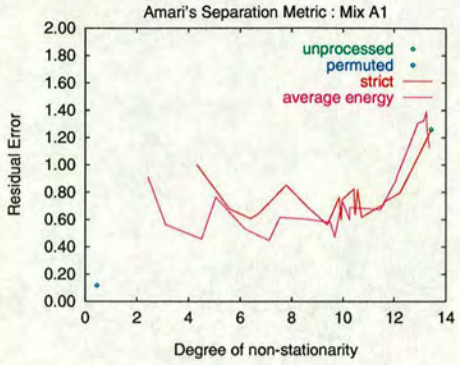
The final performance results were tabulated according to the parameters of the silence removal process used to generate the sources from which the inputs were mixed to facilitate linkage

with the non-stationarity assessments of the processed sources. The pairs of corresponding values (degree of non-stationarity and separation performance) were extracted and re-ordered according to the total degree of non-stationarity of the sources. These values were then graphed for each of the silence identification methods, and are shown in Figure 4.7. It should be noted that in all of the graphs the data are concentrated towards the higher end of the degree of non-stationarity scale, and that the graphs from the energy-based approach all have several points clustered towards the high non-stationarity side. The greatest number of input sets' assessments sum to this area due to the overlap of effect of the different parameters of the non-stationarity reduction. This trend slightly skews the emphasis of the graphs, but they can still be seen to exhibit a roughly 'U'-shaped curve, with their minima around the mean degree of non-stationarity — close to the value 10. The trend is clearer if the extreme left-hand data points of Figures 4.7(b) and 4.7(d) are considered outliers and omitted, due to their excessive disparity from the rest of the data, and the general trend across all of the other graphs. Although the graphs do not show a great difference in performance, their shape suggests that there is a key level of non-stationarity around which separation performance hinges.

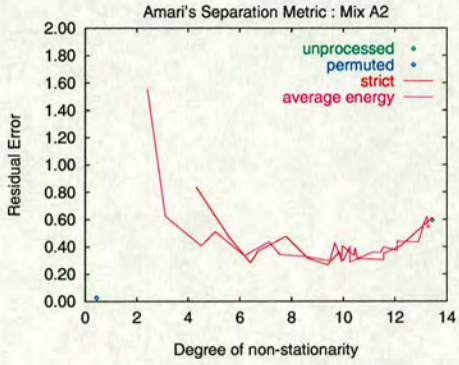
Visually examining the graphs for trends across all of the different mixtures yields the observation that the bias of the 'U'-shaped curve changes with the degree of difficult of the mixing. At the most difficult mixing (Figure 4.7(a)), the bias is towards the left-hand side — the more stationary sources yielding better separation performance. However, as the magnitude of the determinant of the mixing matrix increases, and hence the mixing matrix becomes better conditioned and more easy to solve, the bias shifts through a more central level (Figures 4.7(b) and 4.7(c)), and then increasingly to the right-hand side (Figures 4.7(d), 4.7(e) and 4.7(f)). This shows that for the easier mixings, the more non-stationary sources actually yielded better separation performance. The likely reasons for this will be covered in Section 4.8.

Relationship with the silence removal parameters

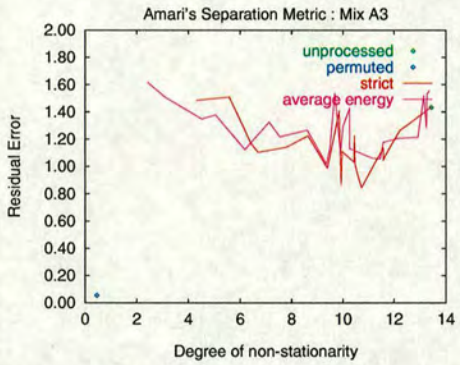
The experiment was designed to facilitate investigation into the relationship between the silence removal process parameters and the separation performance. The separation results were grouped according to the silence removal algorithm employed, and the mixing matrix used in generating the inputs. Each group of results was then graphed over the range of thresholds used at each duration of interest, and over the range of durations at each of the threshold levels. However, the resulting two sets of graphs showed little difference, so only the first set of these



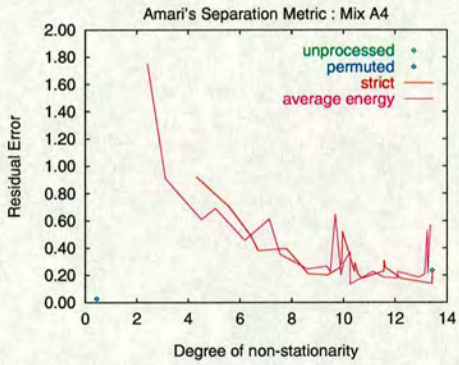
(a) Separation performance against non-stationary assessment : mix **A1**



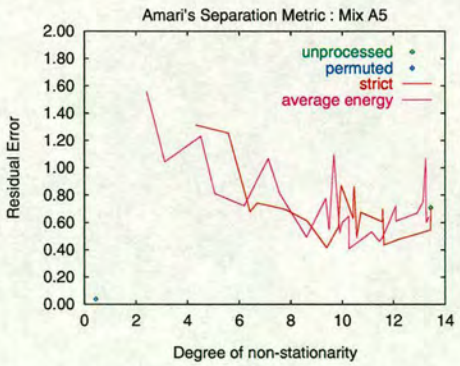
(b) Separation performance against non-stationary assessment : mix **A2**



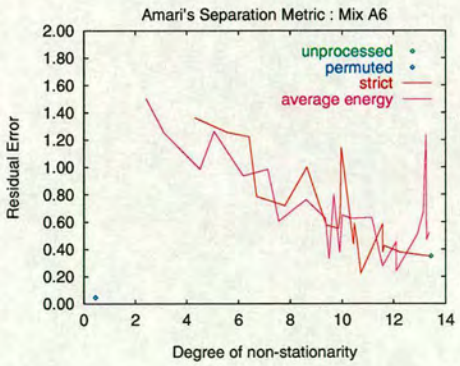
(c) Separation performance against non-stationary assessment : mix **A3**



(d) Separation performance against non-stationary assessment : mix **A4**



(e) Separation performance against non-stationary assessment : mix **A5**



(f) Separation performance against non-stationary assessment : mix **A6**

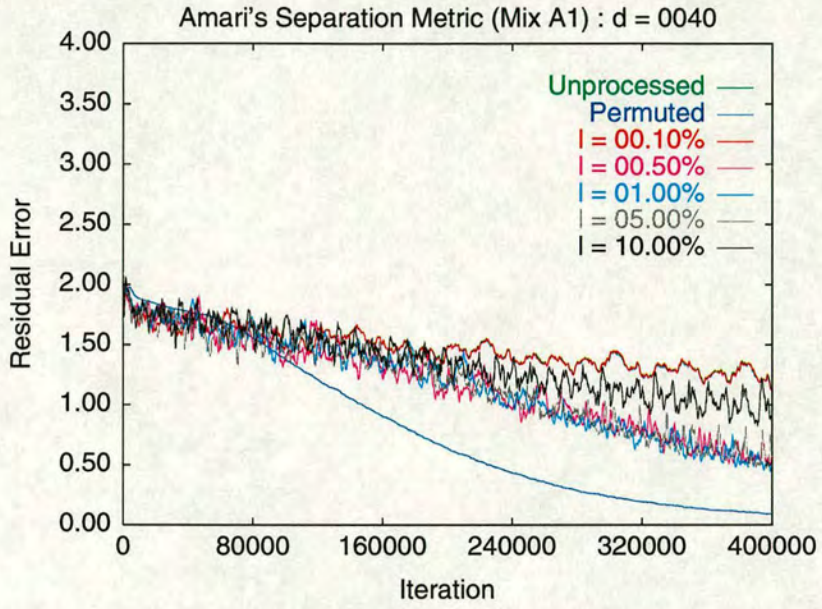
Figure 4.7: *Overall separation performance versus degree of non-stationarity*

(varying thresholds at a fixed duration, for the strict thresholding approach) will be discussed in detail here, in Figures 4.8 and 4.9.

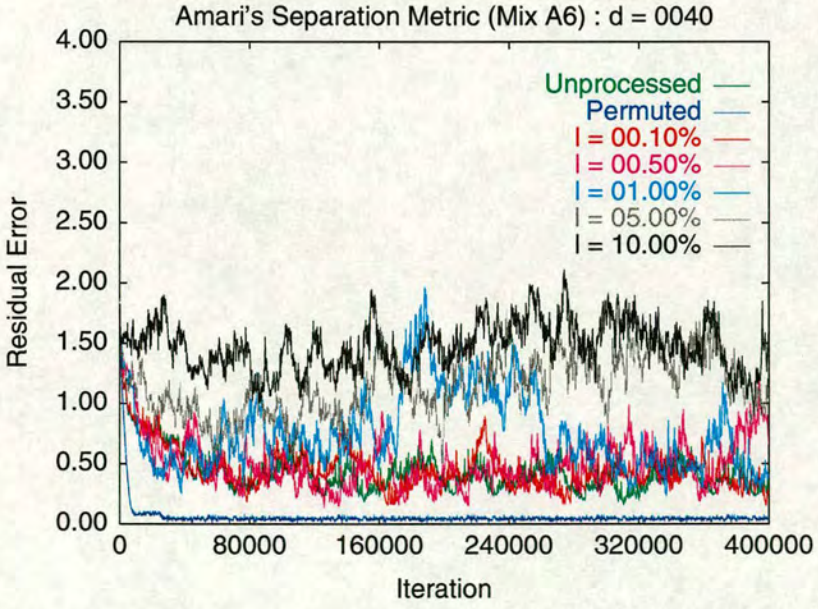
For comparison purposes, the results of the corresponding separations using the unprocessed sources, and those using the permuted sources were also included on each graph, as these should provide bounds within which the experimental results should lie. Graphs of the least well-conditioned mixtures (*i.e.* generated using the mixing matrix with determinant of smallest magnitude, **A1**), and of the most separable mixtures (matrix **A6**) were generated for each of the cases considered. From the analysis, the results of the others lay in between the two, with the exception of mixture **A4** which fared particularly badly. From the graphs, at short durations (Figures 4.8(a) and 4.8(b)), the range of the performance assessment graphs shows greater variation for the more separable mixture than for the other. Mixture **A1** shows improvement in separation performance at both durations, while mixture **A6** suffers a worsening of performance at the shorter durations. For both mixtures shown at the shorter duration, the middle three threshold levels produce the greatest difference in the results, while at the longer durations it is the top three threshold levels that make the most difference.

From the graphs, a precise estimate of ideal parameters cannot be made, but indicate that moderate threshold values of 1.0% or 5.0% give the best results. Similar trends in the graphs of performance over varying durations (Figure 4.12) indicate that relatively long durations of around 0.1 or 0.5 seconds are also beneficial. These observations agree with the previous comments on the relationship between the degree on non-stationarity and the separation performance — that it is the middle range parameters which lead to non-stationarity assessments that are also close to the middle of their range. The most apparent trend across the different mixtures is that the less well-conditioned the mixing matrix, the greater the improvement in performance offered by the use of non-stationarity reduction. In practice, for some cases of the easier mixtures, separation performance of the pre-processed signals is worse than that of the original sources. Statistical analysis on subsequent runs, averaged over all initial weight set configurations and start times (see Appendix B), showed that overall, the differences in separation performance due to the silence removal using the strict threshold assessment at all combinations of the duration and threshold parameters was not significant at the 5% level.

The results from the energy-based thresholding experiments are presented in Figures 4.10 and 4.11. The graphs show a similar spread of results, but with clearer distinction between the

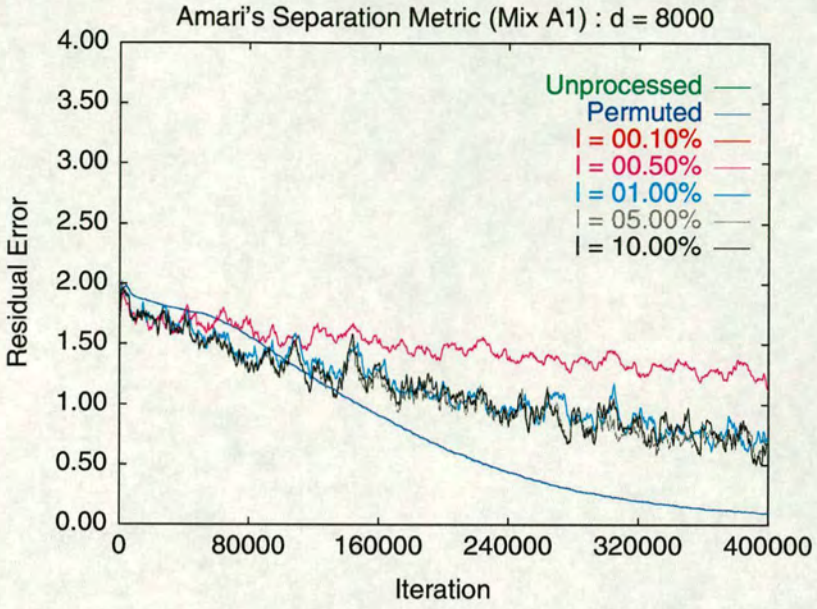


(a) Mix **A1**, for $d = 0.005$ s

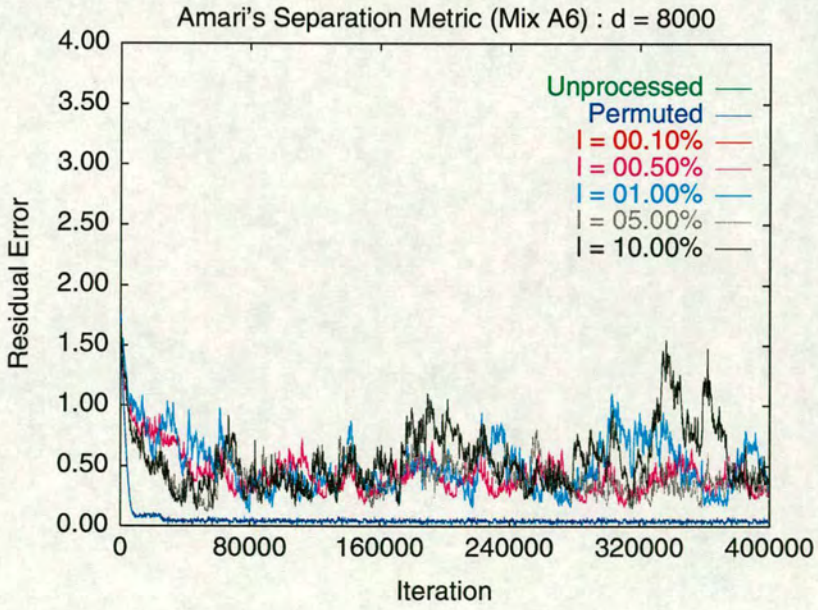


(b) Mix **A6**, for $d = 0.005$ s

Figure 4.8: *Results from the separation of strictly thresholded sources, for fixed duration* $d = 0.005$ *s, and over a range of threshold levels,* l



(a) Mix **A1**, for $d = 1.000$ s



(b) Mix **A6**, for $d = 1.000$ s

Figure 4.9: *Results from the separation of strictly thresholded sources, for fixed duration* $d = 1.000$ *s, and over a range of threshold levels,* l

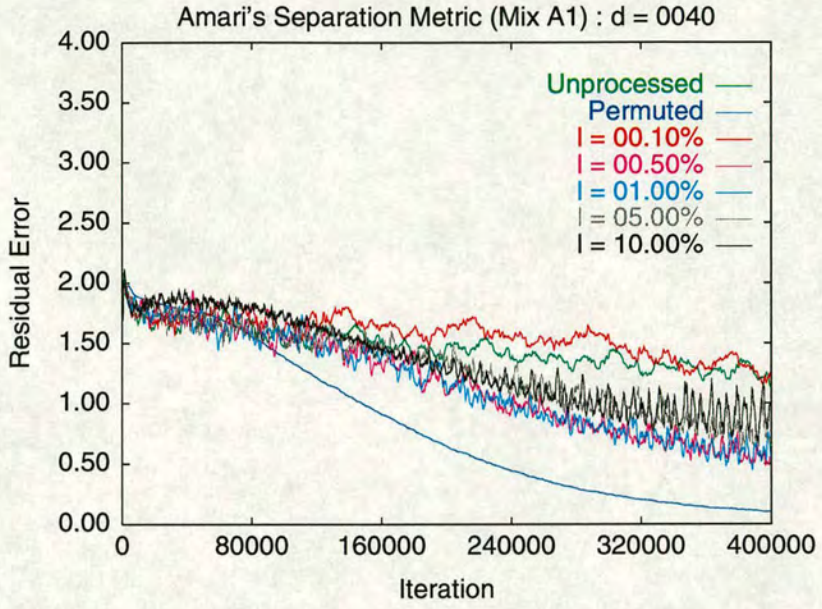
different traces and faster convergence at longer durations for the less well-conditioned mixture. The range of variation in performance was slightly worse for the **A6** mixture than that of the **A1** mixture. Figure 4.12 shows the results from the energy-based approach for a fixed threshold level of 10.00% for varying durations. For the longer durations, the rate of convergence is initially even faster than that of the permuted source mixtures, although the former did not achieve the ultimate degree of separation of the permuted source mixtures. (The permuted source mixtures were not included in the statistical analysis for this experiment.) The fastest convergence at this threshold level can be seen for the duration $d = 0.10$ s.

Performance improvement

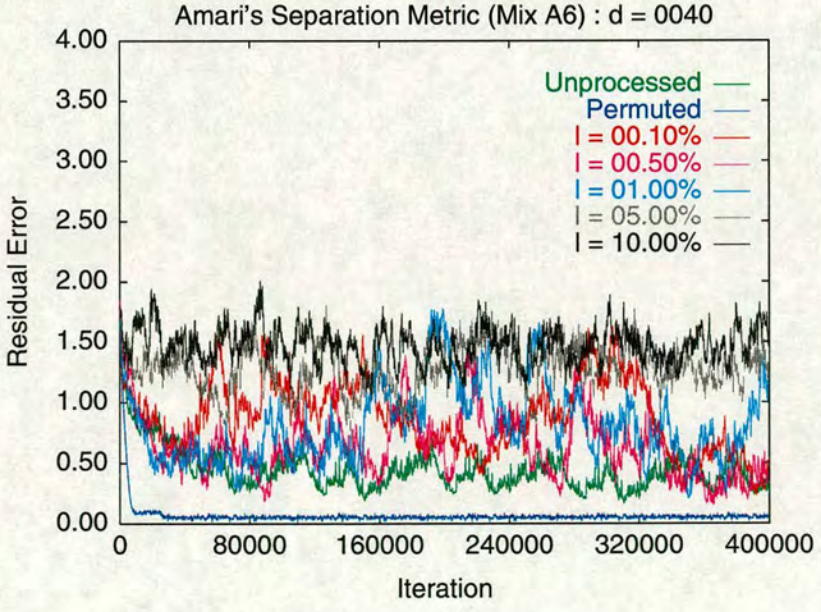
Although the graphs of Figures 4.8, 4.9, 4.10, 4.11 and 4.12 illustrate some improvement, it is hard to quantify this in terms of the quality of the source signals since the metric refers to residual error in the product of the mixing matrix and the network weight set (unmixing matrix). However, Barros & Ohnishi [53] claim that their metric provides more meaningful values that can be attributed to the difference in performance and can be mapped to a percentage indicating the degree of separation of each signal. Using this procedure, the difference in separation performance between the unprocessed case and the cluster of traces shown in Figure 4.13 for the **A1** mixing matrix was estimated to be approximately 10%. From these experiments it can be seen that a greater difference in performance is apparent when dealing with the less well-conditioned mixing matrix, and that durations of 0.10 s and 0.50 s, and threshold levels of 1.00% and 5.00% enhanced rates of convergence in the simulations. However, these variations did not significantly affect the final separation performance. These agree with the conclusions relating the degree of non-stationarity to the separation performance. Further discussion of these results is deferred to Section 4.8.

4.7.4 Batch sizes

Mixtures of signals covering the full range of degree of non-stationarity determined above were created as before, using the hardest and easiest mixing matrices (**A1** and **A6** respectively). These mixtures were passed to networks configured with different batch sizes, from 50 to 500, to allow the effect of non-stationarities incorporated into the batch updates by the summing process to be analysed. The batch sizes selected correspond to intervals of 6.25 ms through to 62.50 ms, which span the stationarity threshold (approximately 20–25 ms). A constant learning

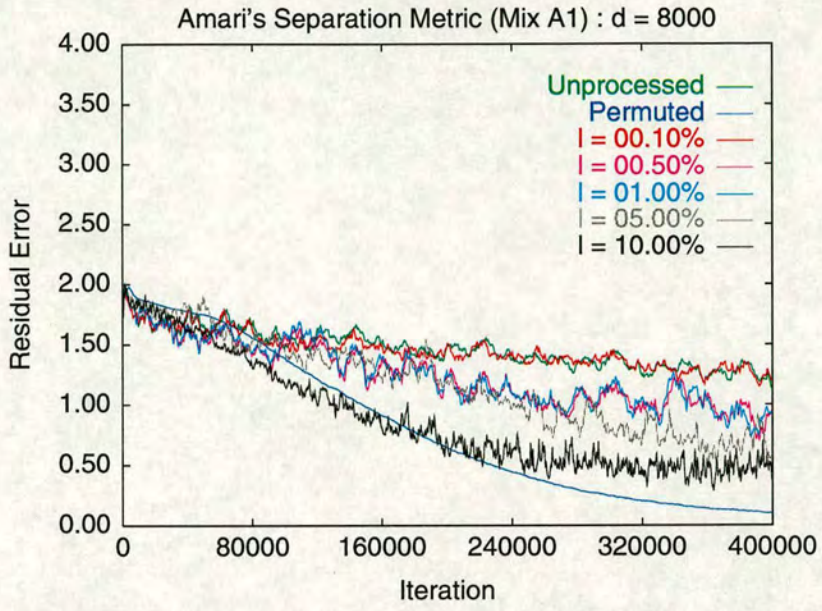


(a) Mix **A1**, for $d = 0.005$ s

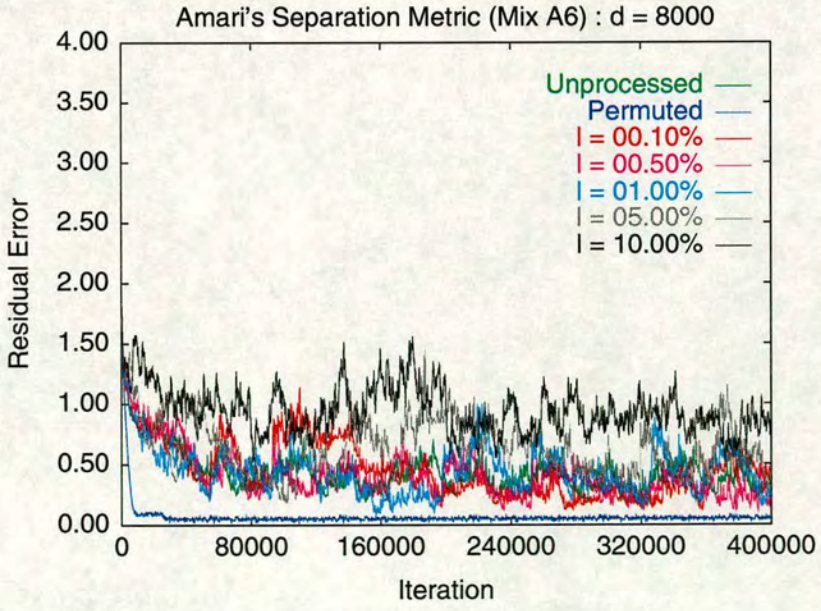


(b) Mix **A6**, for $d = 0.005$ s

Figure 4.10: *Results from the separation of average energy thresholded sources, for fixed duration* $d = 0.005$ *s, and over a range of threshold levels,* l



(a) Mix **A1**, for $d = 1.000$ s



(b) Mix **A6**, for $d = 1.000$ s

Figure 4.11: *Results from the separation of average energy thresholded sources, for fixed durations* $d = 1.000$ *s, and over a range of threshold levels,* l

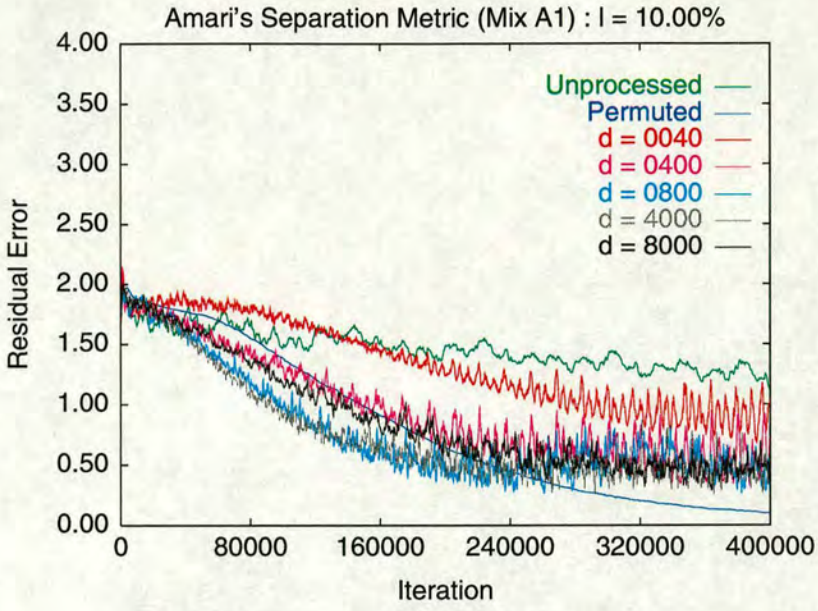
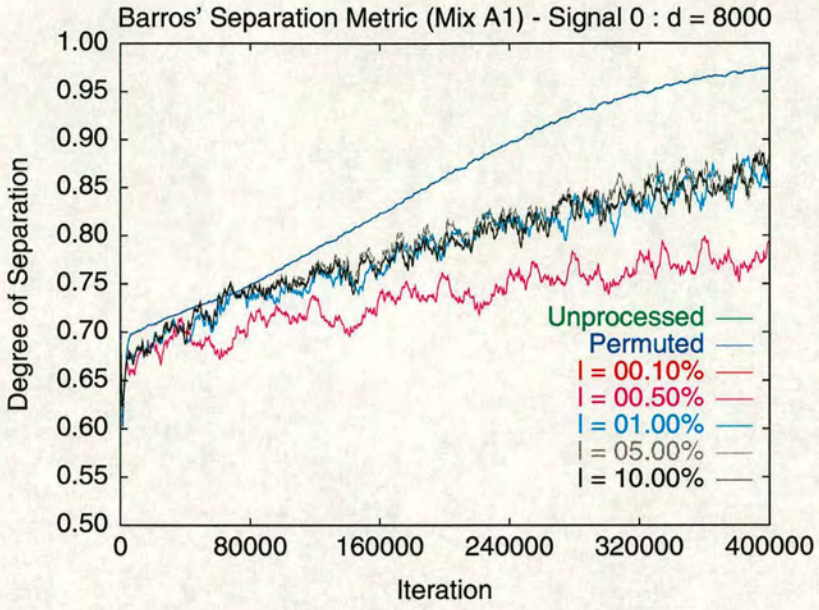


Figure 4.12: *Separation performance for mixture* **A1**, *at fixed threshold level* $l = 10.00\%$, *over a range of durations,* d

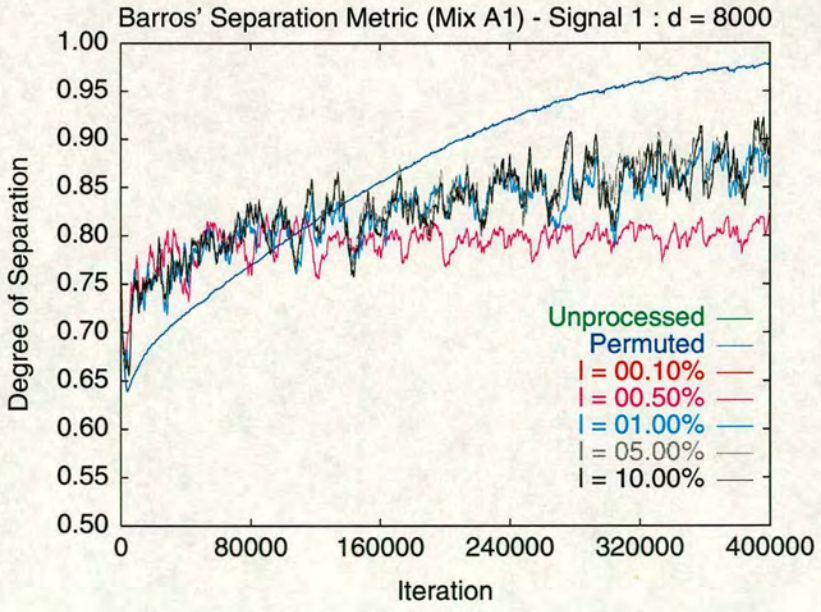
rate of 0.01 was maintained. Selected results from the two different mixtures using the sources of the highest and lowest stationarity are given in Figures 4.14 and 4.15.

For the **A6** mixtures, the graph for low stationarity indicated that the smallest batch size led to the best level of separation performance as shown in Figure 4.15(b). However, statistical analysis (see Appendix B) showed that this was exceptional and in the other cases, batch sizes of 200 iterations and 500 iterations gave significantly better performance (as illustrated in Figure 4.14(b)). Graphs for the **A1** mixtures showed that there was little difference in performance at both high and low stationarity (differences statistically not significant at 5%). The performance graphs show much less variation during the separation, and the difference in performance is in most cases only marginal (see Figure 4.15(a)). The larger batch sizes often resulted in faster convergence towards the ultimate solution, but with a greater variation in value in the process — this is most evident in the tests with signals of higher stationarity (Figure 4.14(a)).

Following the same reasoning as in the previous section for the difference in general performance between mixing matrices, the trend in the batch size results was as expected — for the **A6** (easier) mixture, which can be tracked more easily by the adapting network, the larger batch sizes (with the more stationary updates) produce better results than the smaller

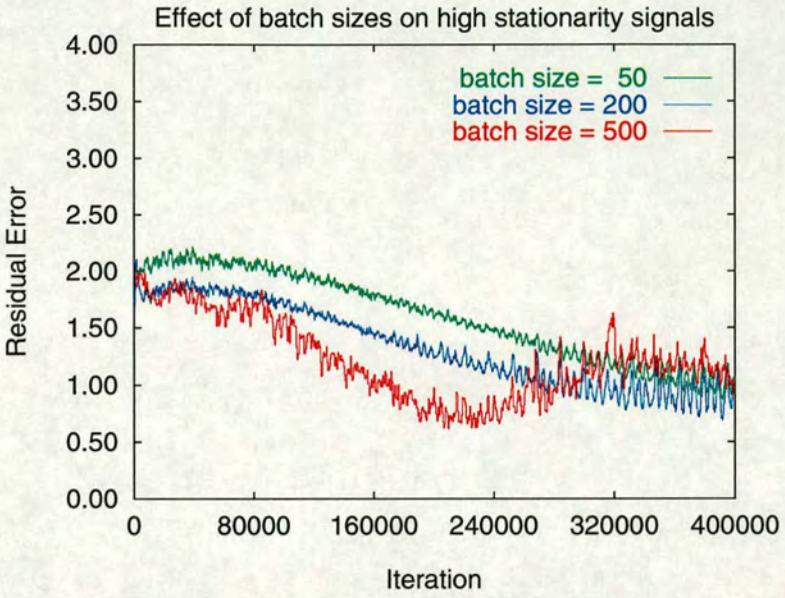


(a) Signal 0 from mix **A1**, for $d = 1.000$ s

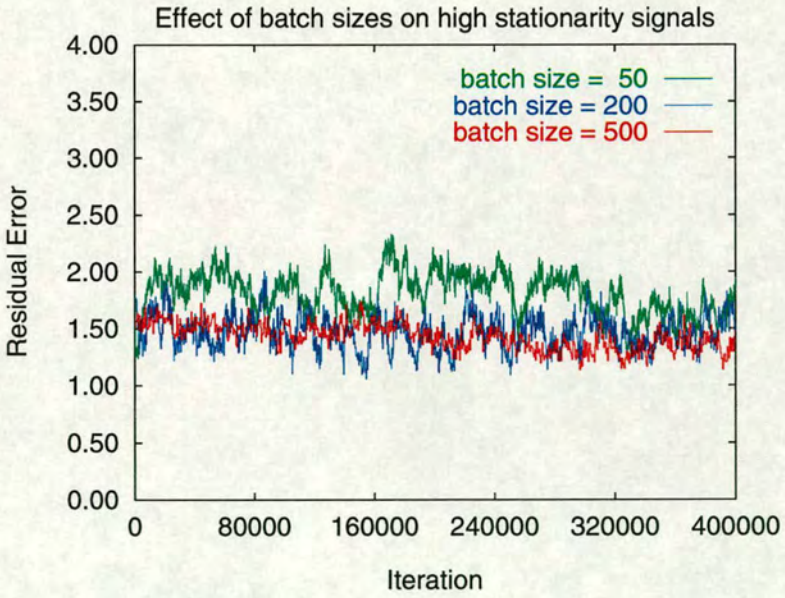


(b) Signal 1 from mix **A1**, for $d = 1.000$ s

Figure 4.13: *Barros and Ohnishi's performance metrics for the* **A1** *mixture at* $d = 1.000$ *s, over a range of threshold values* l

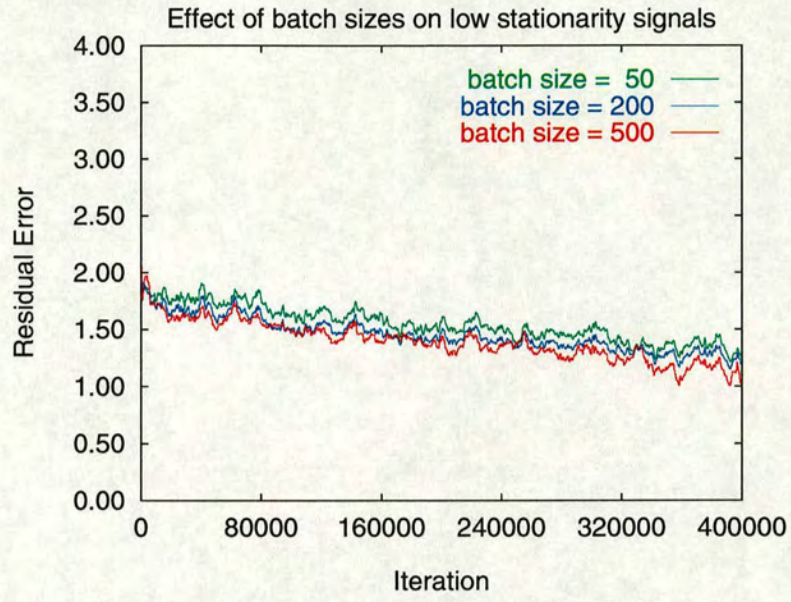


(a) Mix **A1** : high stationarity

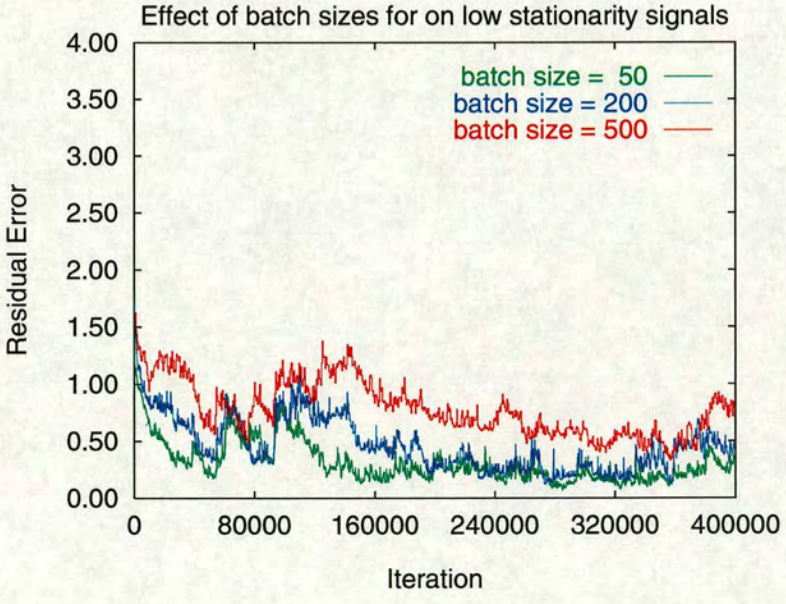


(b) Mix **A6** : high stationarity

Figure 4.14: *The effect of batch size on separation performance for high stationarity*



(a) Mix **A1** : low stationarity



(b) Mix **A6** : low stationarity

Figure 4.15: *The effect of batch size on separation performance for low stationarity*

one. With the **A1** (harder) mixture, where convergence is generally smoother, the network cannot adapt quickly enough to follow all the rapid changes in the source signals. In these cases, the larger batch sizes improves the initial rate of convergence as it moves the weight set further along the convergence trajectory more quickly at any particular update. For more stationary sources, this results in a faster path to the approximate area of the solution, but the system then suffers from the overshoot problem typical of poor batch size / learning rate selection. The less stationary sources result in a generally smaller batch update, since they average out to follow the overall signal trend.

In most cases, and for both sets of mixtures, the results of the tests run with a batch size of 200 samples (the value selected by Bell & Sejnowski, and used in the experiments of the previous section) offers the best performance. This value is likely to be dependent on the characteristics of the source signals, and in this case corresponds to a period of 25 ms ($200 * \frac{1}{8000}$), which is the approximate threshold of stationarity in speech signals.

4.8 Discussion

The results generated from the experiments presented in this chapter address several of the key points of an investigation into the effect of typical speech signal non-stationary characteristics on the separation performance of an information-maximisation-based blind signal separation system.

4.8.1 Non-stationarity assessment

The assessment of non-stationarity is integral to this investigation, and hence some means of quantifying the degree of non-stationarity exhibited by the signals is essential. There are many aspects to be considered in defining such a measure, but the principal one of relevance to this work is the uniformity of the variance of the signals in question. Therefore, the effect of changing magnitude of this characteristic was investigated. The co-efficient of variation (CV) can be used to to measure the constancy of a set of data. The metric created to assess the degree of non-stationarity of the signals uses the CV to measure the change in range of the signal variance over the duration of the signal. Although the figures themselves have no obvious link to features of the signal, they can be related to an observational assessment of the change in signal waveform before and after processing. Since all other aspects of the experiments had

been controlled, these figures can be used as an index of the desired variable, the degree of non-stationarity of the signal(s), against the performance of the separating system.

Other properties that could have been researched are the frequency of the changes and the rate at which these changes occur, *i.e.* how quickly the variance shifts from one level to the next during the changes. Both properties affect the suitability of the information-maximisation-based algorithm to this work, but are more related to the batch sizing comments (Section 4.8.4), due to the temporal aspects that they embody.

4.8.2 Non-stationarity reduction

The principal reason for the selection of silence removal as a means of non-stationarity reduction was its simplicity in achieving its intended purpose. Simplicity was considered important from the point of view of idealised modularity, better interpretation of the results and the possibility of future extensions to the work (such as a hardware implementation of a blind separation system incorporating these modifications). Whilst more complicated techniques do exist, the additional benefits offered at the cost of their complexity fall outside the boundaries of this study.

The effect of the silence removal on the signal waveforms is apparent, and hence a visual inspection of the processed signals allows a basic assessment of the performance of the techniques. Such an inspection also identifies the reduction in non-stationarity of the signals — the envelope of the variance is more uniform after processing. This links well with the non-stationarity assessment metric discussed above.

The parameters of the silence removal process (duration and threshold level) affect the degree of non-stationarity of the processed signals through their direct relationship to the sections removed, and hence the effect of a change in either of them can be anticipated to some degree. The spread of the results of the energy-based variant was closer to that expected than those of the strict method, due to its improved noise tolerance. This variant also generally results in more of the signal being removed, particularly at the beginning and end of inter-word gaps, where there is a marked change in signal level.

Of the two parameters, the threshold level plays the more critical rôle in the amount of signal rejected. Variation of the threshold produced a wider range of values at a fixed duration than variation of the duration at a fixed threshold. Subsequent statistical analyses indicated that

duration had no significant interaction with the other experimental factors. Within the range of values tested, no lower limit on the reduction of non-stationarity was found, although the quality of the signal becomes progressively worse as more is removed. Optimal values for the two parameters were found to be in the middle to upper regions of the ranges considered — around 0.1 to 0.5 seconds in duration, which corresponds to the inter-word gaps found in speech signals, and roughly 1.0 to 5.0% for threshold level.

The durations identified are well above the stationarity threshold for speech signals, and also above the optimal batch size found as well. Hence the sections removed were definitely contributing to the non-stationary of the signal variance over the period in question. Also, at these longer periods the average noise level of the sections considered is lower, allowing a lower threshold to be set, at least for the energy-based variant. The relatively low optimal thresholds found are in accordance with this, since there must be a minimum threshold level which allows the elimination of the desired period of inactivity. However, there is nothing to be gained by removing more of the signal above this level.

While these results were more or less as anticipated, the trend of the separation performance results corresponding to the degree of non-stationarity was not.

4.8.3 Blind separation performance

Initially, it would seem that since there is no active signal present from which the system can extract information during periods of silence, that the removal of such 'reduced information' periods should not detrimentally affect the separation performance. In fact, since the weights of the separating network have been shown to diverge during these periods, their removal should result in an overall improvement in performance over a given period of time. However, looking at the system as a whole, rather than individual signals, such periods of inactivity in one source may allow more information about one of the others to be inferred. Consequently there must be a trade-off between the two views.

Signal characteristics

In this experimentation, no simple relationship between the degree of non-stationarity and the separation performance was found. Initially, it had been thought that there would be a linear relationship between the assessment metric and the separation performance, due to the

sensitivity of the algorithm to the signals' variance. The 'U'-shaped curves were not anticipated and imply some optimal level of non-stationarity above the minimum. Due to the approaches used in the non-stationarity assessment and reduction processes, it was not possible to reverse the mapping and determine the optimal silence removal parameters from the optimal measures of non-stationarity — different combinations of duration and threshold values could all lead to the same degree of non-stationarity.

This relationship can be explained by considering the effect of the changing signal characteristics on the update rule. Since the rule is attempting to converge to a solution based upon these characteristics, changes in the characteristics, such as those that occur between periods of activity and inactivity in speech, will hinder the system in reaching a stable solution. The system will attempt to track the solution defined by the current characteristics, but depending on the magnitude and frequency of the changes, may never fully reach it. Even if it does reach a solution, the system may have to re-adjust at the next such change. Some small degree of non-stationarity may prove beneficial as it will mean that the surface that the gradient-ascent rule is attempting to scale will be continually changing. Such changes may assist the system to escape from any local maxima it encounters, since they may not be permanent features of the correspondingly changing surface.

Conditioning

The conditioning of the mixing matrices also appeared to affect the relationship between the degree of non-stationarity and the separation performance. The lowest point of the 'U'-shaped curves (*i.e.* the point of best performance) was shifted towards the more stationary combinations for the less well-conditioned mixtures. For such mixtures, even relatively small changes in pivotal values derived from the source characteristics could result in large changes in the network's position in solution space. Hence, minimal changes in the characteristics are desirable to improve the algorithm's chances of converging. If any such changes do occur, the system will take a long time to start to converge to the new solution. If the solution returns to (near) its former location after only a short period of time, the weight set may not have had a chance to diverge too far from its earlier position. Since the typical changes in variance of a speech signal are likely to be centred around two main loci (one for the active parts of the speech, and one for the non-active parts) such changes may not be too damaging if the time spent converging to the non-active solution is minimised. Therefore, it could be expected that

the removal of the periods of silence that define this alternative convergence target would assist in this respect. However, this was not supported by the statistical analysis of the experimental data.

For the well-conditioned mixtures, convergence is comparatively fast and even during the relatively short time that the system spends at the non-active target, the network's weight set changes considerably, moving away from the desired solution. This would explain the poorer separation performance seen for these well-conditioned mixtures.

Noise

Finally, the traces of some of the performance metrics generated for the results of the non-stationarity reduction experiments were seen to be rather noisy when compared to the smooth traces of the "permuted sources" experiments. This is due to a combination of the bursty nature of the speech and the batch size selected, causing the network to oscillate around its true solution when it is unable to generate a sufficiently small update to allow it to reach the peak of the constraint surface. This is a common problem in gradient-based learning situations, and there are techniques available for overcoming it. Simulated annealing, for example, could be used to reduce the size of the update by lowering the learning rate. Reducing the batch size, and hence the number of updates summed together, should achieve a similar effect and allow the system to settle to a more accurate solution.

4.8.4 Batch sizing

Batch sizing in the experiments is more than a computational optimisation when the temporal aspects of the situation are taken into account. As well as influencing the accuracy of solutions reached, the size of the batches also has relevance to the non-stationarity of the signals, and the period of time over which the batches are amassed. The results from this study provide a rationale for the selection of an appropriate batch size (200 iterations, in these experiments) which concurs with that reported in Bell & Sejnowski's original paper [10]. For the 8 kHz signals considered in both cases, this corresponds to a 25 ms batch, the upper limit of stationarity for speech signals. This seems justifiable since below this level, the signals (and hence the updates in the batches) will be stationary, and therefore the longest batches possible should be used. This period allows for absorption of fluctuations during the active parts of the

speech. Beyond this threshold, however, the non-stationarities of the signals may begin to affect the batch updates generated, and consequently impact on the algorithm's performance. This is especially true of inter-word gaps in the speech, which must be tracked.

In light of this, it would also be interesting to relate the non-stationarity assessment / reduction to batch size and look for correlations, although the sequential nature of the batches may not be appropriate given the continuous assessment and reduction processes. These effects are linked to the signal characteristics, and hence the values reported here can only be used as a guide for other speech signals — other types of signal would required similar analyses. As noted earlier, other aspects of non-stationarity (such as rate and frequency of change) that could be investigated would also be affected by batch-sizing, and therefore should be taken into consideration in the design of future experiments.

4.8.5 Statistical analysis

The statistical analysis did not demonstrate significant differences between the performance achieved by the experimental and original algorithms, at the 5% level. The main difference between the techniques was the apparent speed of convergence rather than the separation performance *per se*. In several instances, interactions between the various factors of the experiments were shown to be significant, but when these were graphed there was no consistent trend or simple interpretation of these interactions. The statistical significance of these interactions was due to the high number of replicates and the the high precision of the data, but in many cases they do not appear to be scientifically significant. As expected, the effect of the different mixes was found to be significant in most cases, whereas the duration parameter was shown to have no significant interaction.

4.9 Areas for further investigation

The development of the techniques presented in this chapter into an on-line algorithm, and the investigation of its performance, is the subject of Chapter 5. Other areas of study that could be further investigated, leading on from this work are listed below.

- Development of entropy-based criteria, or other techniques, for silence identification and removal which would better align with the framework of the information-maximisation

approach.

- Performance on artificial data, to allow stronger conclusions to be drawn. The parameters of the non-stationarity, such as amplitude and rate of change of the variance, could then be accurately controlled to assess critical levels of each, with regard to the separation performance.

These proposals may not only be relevant in the context of extending the work of this thesis, but would also form useful pieces of research in the field of blind signal separation or independent component analysis.

4.10 Summary

This chapter has examined the poor performance of the information-maximisation blind signal separation algorithm on unprocessed speech signals, which exhibit long-term non-stationary characteristics. A method of quantifying the degree of non-stationarity was proposed. Two variations of a silence assessment technique that can be used in silence removal to change the degree of non-stationarity of such signals were examined — a strict, consecutive-sample threshold approach, and a buffered, energy-based method. Appropriate values for the range of parameters to be used for the silence identification were determined to lie between 0.10 seconds and 0.50 seconds, and between 1.00% and 5.00%. Initial observation of the graphs indicated that combinations of parameters that resulted in a medium degree of non-stationarity gave the best separation improvement. However, statistical analysis of the data showed no significant difference.

The relationship between the total degree of non-stationarity of the sources and the separation and convergence performance of the network weight set in an infomax-based blind separation system were illustrated for signal mixtures of varying complexity. This was found to be a 'U'-shaped curve, skewed to one side or the other according to the magnitude of the determinant of the mixing matrix used. More well-conditioned mixings biased the curve towards the higher end of the non-stationarity scale. This is because the presence of non-stationarity can assist the well-conditioned systems in escaping from local extrema, while the pivotal effect of ill-conditioned systems means that this is unnecessary. In these cases, non-stationarity is detrimental as is slows the rate of convergence to the true solution by drawing the convergence

away to new, short-term targets. The separation performance varies more widely over the course of the simulations for the more stationary sources.

The selection of batch sizes for use in the update algorithm was shown to affect significantly the final performance level achieved after a fixed number of iterations, especially for inputs created from more stationary signals and mixing matrices with determinants of smaller magnitude. A value of 200 iterations per batch, corresponding to a time period of 25 ms, on the borderline of stationarity for speech signals was found to be optimal.

Chapter 5
Non-Stationarity Reduction in an Adaptive, On-Line Blind Signal Separation System

Chapter 4 investigated the effects of source signal non-stationarity reduction by silence removal on the separation performance of an information-maximisation-based blind signal separation system, when this reduction was carried out prior to mixing. This is not possible in a real-world blind separation problem, where there is no *a priori* knowledge of the source signals. This chapter investigates the incorporation of these non-stationarity reduction techniques into an on-line adaptive system. Additional strategies for further improvement of the separation performance are evaluated, and performance comparisons are made with other existing approaches to the blind separation of non-stationary sources.

Although some aspects of this work have been studied by others, including Van Gerven, Van Compernolle, Nguyen and Jutten [6–9] under the name of intermittent adaptation or non-permanent learning, this chapter investigates and evaluates the extension of the technique proposed in Chapter 4 — that of silence removal.

5.1 Outline of the investigation

Situations where pre-processing is not possible were considered, along with strategies for tackling this. In a truly blind situation, the source signals are unknown and hence cannot be pre-processed in any way. However, since the outputs of the separating network are estimates of the source signals, these can provide information about the source characteristics. For a real-world application, all of the processing must be done in an on-line, if not in a real-time, manner. The potential of such applications is high, particularly given their suitability for processing speech signals and the number of potential communications-related problems to which these techniques could be applied. The definitions of the terms "on-line" and "real-time" are given here for clarification :

On-line systems On-line systems process data in such a manner that the continual stream of input data need not be halted for the processing to take place. An arbitrarily large, but finite, delay may occur between the input of a specific item of data and an output dependent on that input, given sufficient buffering.

Real-time systems Real-time systems operate in a manner and at a speed that produces outputs at the same rate as that of the arrival of the inputs. The time lag between the arrival of the inputs and production of outputs derived from those inputs is smaller than that of successive input arrivals, so that the results are instantaneously available.

An on-line separation system should not miss any changes in the mixture of the signals, which could otherwise lead to poor separation performance. Such changes could be due to variation in the signal characteristics, or in those of the mixing environment. The ability to compensate for such changes, and maintain separation performance is known as "tracking" [70, 83]. An advantage of investigating an on-line *audio* separation system is the ability to assess the quality of the outputs, by listening to them. For this, a fixed sample rate must be maintained at the outputs. In the experiments performed in this study, the output format was nominally chosen to match that of the inputs — a raw 8 kHz sampled signal.

In addition to the modification of the information-maximisation separation system to incorporate the non-stationarity reduction methods, additional variations based around on-line adaptive neural update strategies were also examined to see if they offered worthwhile performance improvements. Comparisons between the performance of these methods, and other existing separation techniques were carried out, to assess the merits of each approach.

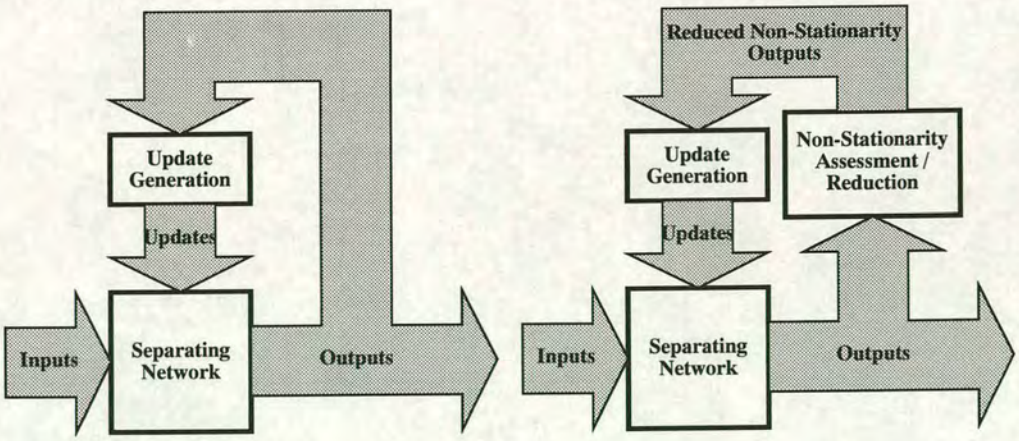
5.2 Modification of the infomax-based system

To use the techniques proposed previously for dealing with non-stationary source signals (Section 4.3.2), certain modifications had to be made to the experimental setup. Firstly, the non-stationarity reduction processing had to be incorporated into the system architecture to enable the processing to be carried out in an on-line manner. Furthermore, since there is now no direct access to the source data, the same form of pre-processing that was used in the experiments described in Section 4.6.1 can no longer be carried out. Applying the techniques to the inputs of the system instead is not practical, since it has been illustrated in Section 4.6.1 that it is difficult to identify when one or more of their constituent sources has fallen silent. The

reason for the difficulty is that the inputs may contain mixtures of unknown numbers of active sources at different times and cannot be assessed for the silence of one of more of these without knowing how many signals are present — a non-trivial problem in itself. This non-stationarity reduction of the inputs would not necessarily be advantageous, since the input signals could have a (near-)stationary variance while the sources themselves had considerably non-stationary variances.

An alternative solution to the problem exits. Since the outputs of the system approximate the source signals, the non-stationarity assessment techniques (the silence identification, used previously) can be applied to these outputs instead. Whilst these estimates may not be particularly accurate at the start of the separation run, the earlier experiments showed that they improve quite quickly as the network converges to a separating solution, the rate of convergence being accelerated by the improved update algorithm.

The non-stationarity reduction techniques used in Chapter 4 removed parts of the source signals prior to their mixing and subsequent presentation to the network. In the present study this is no longer possible as the signals being assessed are already the network's outputs. Each set of output samples will have been generated by a set of input samples. However, not all of the inputs need to be used in the calculation of the update for the network weight set, as this calculation can be considered to be on a distinct data path (*i.e.* the feedback loop) from the output generation process (see Figure 5.1). Since it is these updates to the weight set that



(a) The original system

(b) The proposed new system

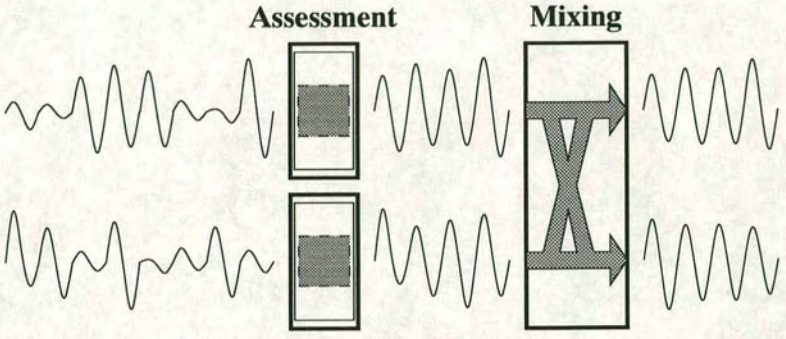
Figure 5.1: *Data-flow for a blind signal separation system*

have the critical effect on the convergence, and consequently the separation performance of the system, the focus of the non-stationarity reduction can instead be placed on this branch of the data path, without affecting the throughput of the outputs.

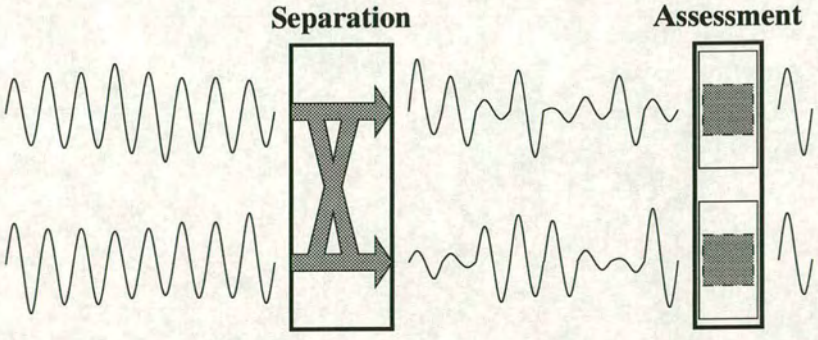
Moving the non-stationarity reduction to this part of the network leads to a significant change in the overview of the processing being done, for this type of system. The source estimates being assessed can no longer be considered in isolation — it is not possible simply to omit one or more of these estimates that fail the stationarity assessment from the update generation, since a full set of input values is required each time for the update algorithm under consideration to process. Hence, if any are to be omitted, all must be omitted, as it is impossible to determine which output corresponds to which source — thus no set of updates can be generated for that set of inputs. This process should still produce the desired effect, since the weight set is protected from sudden large changes in value, which may otherwise prevent ill-conditioned systems from converging, or even cause them to oscillate.

If the weight set updates were calculated in a different way — even using the same basic rule, but derived differently, for individual weights, as is presented in Bell & Sejnowski's paper [10] — it would be possible to update only the weights unaffected by the outputs failing the silence assessment. This is the approach taken by the "intermittent adaptation" described by Van Gerven and Van Compernolle in [6–8], although theirs is based on a simple energy estimate at the outputs of their Symmetric Adaptive Decorrelator for convolved signals. It also makes use of a common absolute threshold as well as relative ones for each channel. Nguyen Thi & Jutten [9] also comment on their "non-permanent learning" technique, which follows a similar strategy, again applied to convolutive mixtures. Further comment on these results is made in the discussion and conclusions in Chapter 6.

The change in the processing focus will result in the performance assessments differing from those obtained during the previous experiments, even for identical non-stationarity reduction parameters. The inputs will no longer have been generated from processed sources, and consequently will have different characteristics. It is possible that none of the new sets of estimated outputs would satisfy the stationarity assessment criteria, and thus no updates would ever being generated (see Figure 5.2) although this is highly unlikely, as long as the parameters selected are within reasonable bounds. Consequently, the development of the separating network weight set will follow a distinctly different convergence route than that of the system operating on the pre-processed sources — but the principles being used to control the update



(a) Pre-processing source signals



(b) Processing source estimates

Figure 5.2: *An illustration of the differences between the signal assessments*

generation are very similar.

All of these additional requirements of the modifications could still be fitted into the block indicated in Figure 5.1(b), between the original outputs and the update generation part of the system. Therefore, no modifications to the learning rules or algorithm were required, other than to allow for the fact that an update may not be generated every iteration.

When used with the batch processing methods mentioned previously (Section 2.2.3.1), the iterations where no update is generated were still included in the batch size count. This meant that when a separation trial was run for the same number of iterations as in the experiments in the previous chapter, with the same batch size constraint, the same number of batches would be generated, even if some of those batches were composed of different numbers of updates.

5.2.1 On-line assessment of signal non-stationarity

Due to the better performance of the energy-based non-stationarity assessment system in the previous sets of experiments, this system was selected for use in all of the on-line experiments.

The use of any non-stationarity assessment in an on-line system could potentially hinder the convergence of a solution at its early stages, when the estimates of periods of silence within the source signals may be unreliable. Consideration was therefore given to enabling this feature only after the network has achieved a certain level of separation. This would still allow improvement in the overall performance to a level that would not otherwise be attainable using the basic separation algorithm alone. Determination of the start for the assessment techniques could be achieved by applying a confidence measure on the accuracy of the outputs. The outputs would initially be poor and improve with time, thereby giving greater weight to the non-stationarity assessment as the simulation progressed, or by a heuristic based on the elapsed separation time. However, this delayed starting was determined not to be crucial to the investigation and hence was not further pursued.

Scaling of the outputs must also be considered in the use of this on-line assessment. Since this is one of the unconstrained features of the problem (permutation of the output order being the other), the thresholding techniques used previously had to be further modified to take account of this indeterminacy. This was achieved by individually scaling the threshold level of each channel by the maximum value seen so far on that channel.

5.2.2 On-line non-stationarity reduction and separation

Use of the modified non-stationarity assessment and reduction techniques used previously meant that the modular block identified in Figure 5.1(b) had to be parameterised by the threshold level and the duration of the period of silence to be identified. In fact, this modularity facilitated their incorporation into the separation system, with only one notable change required — this being the individual scaling of each channel's threshold level (described in Section 5.2.1), to take account of the indeterminate scaling of the output signals. Whilst this approach may not be ideal, as it can be disrupted by a spurious spike in an output of otherwise small magnitude, it can be periodically updated, by simply resetting the maximum to some notionally small value, and then re-assessing it.

No other major changes were required or made to the assessment techniques or to the separation system, other than to make the threshold and duration parameters part of the update algorithm. This change facilitated updating of the parameters by the system during the separation run, if so desired.

A series of experiments were designed using the new algorithm and inputs generated from the original unprocessed, source signals to demonstrate the effect of these modifications. The experiments were run with the same range of parameter values and the same performance assessment metric as used in Chapter 4. The same mixing matrices were used to create the input signals so that the performance results could be compared. Although the earlier experiments had revealed the level of separation attainable after a fixed number of iterations, it was considered appropriate to assess the on-line performance at the final value of the converged separation solution.

5.2.3 Convergence profiling

In order to determine when the system had converged, for each of the different mixing matrices under consideration, a set of experiments were set up to run the simulation for varying numbers of iterations. Thirty runs of the simulation were conducted at each of the different iteration lengths, to provide sufficient data on which to base the assessment. The values of the performance metric at the end of these runs were gathered and analysed, both by plotting them as histograms (not included here) showing the distribution of the values for each mixing matrix, and by evaluating the co-efficient of variation for each of these sets of results.

The histograms showed well-defined peaks around the final performance value when the simulations had converged, with fewer results lying outwith this region. Observation of these distributions were used to guide the selection of the data set on which to perform the further analysis using the co-efficient of variation.

The co-efficient of variation, described previously by Equation 4.5 in Section 4.3.1 can also be used to indicate equivalent levels of convergence for processes with a relatively wide dynamic range, such as the performance metrics for the different mixtures considered in this study. It is calculated as :

$$CV = \frac{\sigma}{\mu} \tag{5.1}$$

Since the performance metric used ideally converges to zero, it would have been possible to use the variance of the results instead, as division by small valued means can lead to misleadingly large values for the co-efficient of variation. However, as some of the mixing matrices used did not yield low values of this metric even at convergence, it was deemed appropriate to retain the division by the mean [120].

To determine the required length of the simulations for a particular mixing matrix, the lowest number of iterations that gave a co-efficient of variation of less that 0.05 was selected.

The results of this assessment for the six mixing matrices used with the original (un-modified) infomax algorithm are presented below in Table 5.1.

$\mathbf{A1} =$	4800000	$\mathbf{A2} =$	3520000
$\mathbf{A3} =$	1040000	$\mathbf{A4} =$	4800000
$\mathbf{A5} =$	2480000	$\mathbf{A6} =$	720000

Table 5.1: *Number of iterations to convergence*

The results of this analysis indicate a high number of iterations are required, equivalent to fifty passes through the data for the longest runs (mixing matrices $\mathbf{A1}$ and $\mathbf{A4}$). These values are far higher than the 480000 iterations used in the original experiments here and those in Chapter 4 and are probably due to the 0.05 convergence requirement, and the near zero means for some

of the runs. These values were used for gathering all data for the statistical analysis.

5.2.4 On-line non-stationarity assessment

In this set of experiments, it is difficult to link performance to a fixed degree of non-stationarity. No single-valued measurement can be made for the processed signals, since they are continually changing throughout the experiment, due to the evolving separating matrix — hence any such measure would also vary during the experiment. Furthermore, it is not possible to use the values determined from the previous experiments, since the signals being processed are different, due to the lack of processing prior to their mixing. Despite this, the duration and threshold parameters provide the closest means of reference to the degree of non-stationarity, as their off-line results and effects are known. Consequently, the experiments described here will use these parameter values which can then be linked back to the effect on the degree of signal non-stationarity previously noted.

5.2.5 Results and discussion

The graphs of results presented provide an indication of the relative performances of the different methods, but due to the noisy nature of the data and the effect of the various start points and times, these must be considered in conjunction with the statistical analyses carried out for each set of experiments. A typical performance graph from a single separation run using the different algorithms, illustrated in Figure 5.3, shows that the separation performance decreases in a manner similar to that previously reported in Section 4.7. However, as can be seen the performance metric levels out at various points where the original algorithm's performance varies considerably during the separation. These points correspond to periods within the output signals (the derived estimates of the source signals) that have been deemed silent. The lack of variation in the performance signal at these points is due to the weight set not being updated, and results in the overall performance more closely following that of the ideal (indicated by the graph for the permuted sources).

It should be noted in Figure 5.3 that there is a large variation towards the end of the separation run — this is due to the output signals satisfying the assessment criterion used. As a result, the algorithm does not suspend the updates, and tracks the variation that occurs at this point while the weights are temporarily thrown from the ideal convergence. Overall, however, general

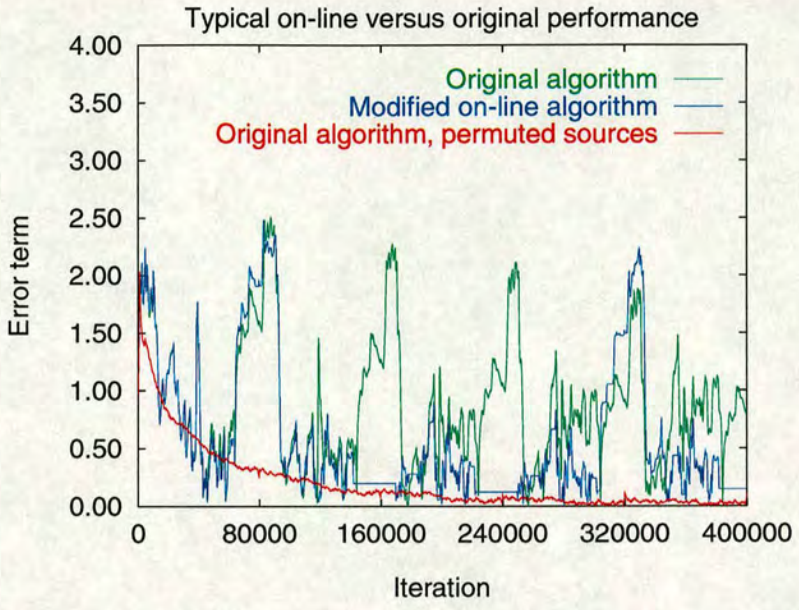


Figure 5.3: *Typical on-line performance*

performance of the new modified algorithm is better than that of the original.

Averaged graphs were viewed to identify trends in the systems' performance and statistical analysis undertaken on the results of the simulations. Although the graphs identified trends, the analyses showed that there was no significant improvement in performance.

From the graphed averaged results and the statistical analysis, the following observations could be made :

Conditioning General performance increased for the better conditioned matrices (with the exception of **A4**), mirroring the trend seen in the results from the unmodified algorithm. However, in all but mixtures **A1** and **A6** the performance was worse.

As before, the original algorithm is better able to track the more well-conditioned mixtures, and hence responds rapidly to the changes caused by the non-stationarity.

Threshold level Generally, the lower threshold levels of around 0.50% and 1.00% produced the best performance levels for the less well conditioned cases. This trend drifts as the conditioning improves, such that for the most well-conditioned matrices, thresholds of 0.10% and 0.50% gave the best results. At the higher threshold levels of 5.00% and 10.00%, performance was frequently worse than that of the original algorithm.

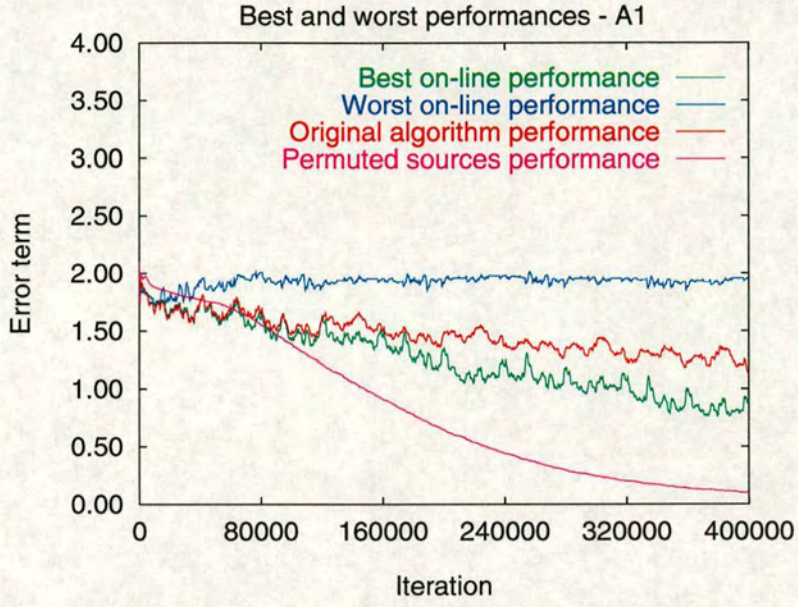
Duration The statistical analysis showed that the interaction of duration and the other factors did not result in any significant difference in performance. However, the duration of the periods of silence being removed affected the rate of convergence of the system, according to the conditioning of the mixing matrix — for the more well-conditioned systems, the longer the periods of silence removed, the faster the rate of convergence. The converse was true for the less well-conditioned cases.

Performance for cases where the periods of silence removed were extremely short varied considerably.

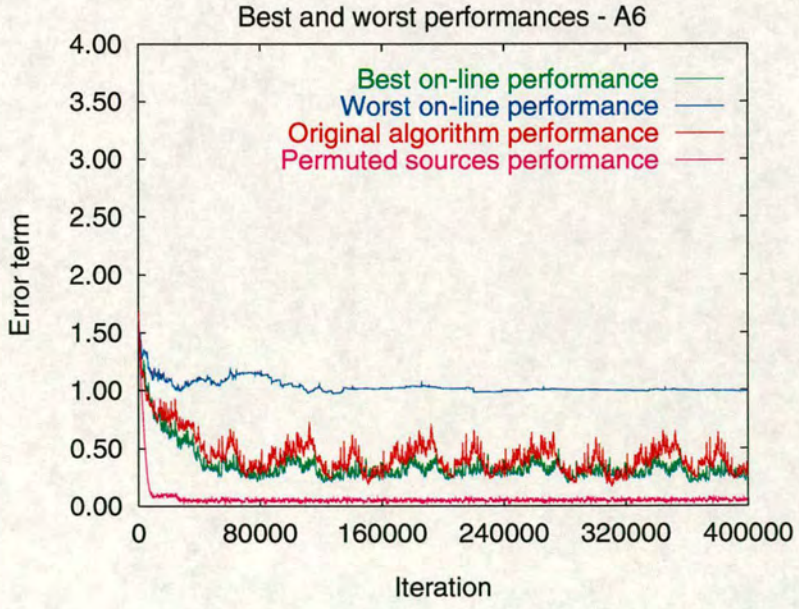
In these tests, the principal characteristic of the performance measures are that they show considerably less variation than those of the original experiments. Examples of the best and worst performances for the least and most well-conditioned mixing matrices are given in Figure 5.4.

5.3 Comparison with the off-line experiments

As noted in the previous section, the general trend with regard to the condition of the mixing matrix is the same as that for the off-line pre-processing — the algorithm offers better performance for the more well-conditioned matrices. The on-line algorithm's performance depends on the accuracy of the source estimation at the network's outputs and as the original, underlying network performs better on more well-conditioned mixing matrices, it is able to converge more quickly to the different solutions caused by the changing characteristics of the non-stationary signals as they change from a period of silence (or at least inactivity) to a burst of speech. Since the signals can be readily tracked, any improvements offered by additional techniques would need to make a big difference to the updates calculated before they would be seen as significant. In the more poorly-conditioned case, finding the solution is more difficult due to the inherent instability of such systems, where even a small change in the updates generated can have a pivotal effect at the outputs. As a consequence, the reduction of the non-stationarities would result in a more noticeable change in separation performance. This is illustrated by Figure 5.5, which shows the best performance of the original, unmodified algorithm on both the processed and unprocessed input signals, on the permuted sources (as an indication of the best performance achievable), and also that of the on-line version of the algorithm on the unprocessed inputs. For the least well-conditioned case, only at durations of

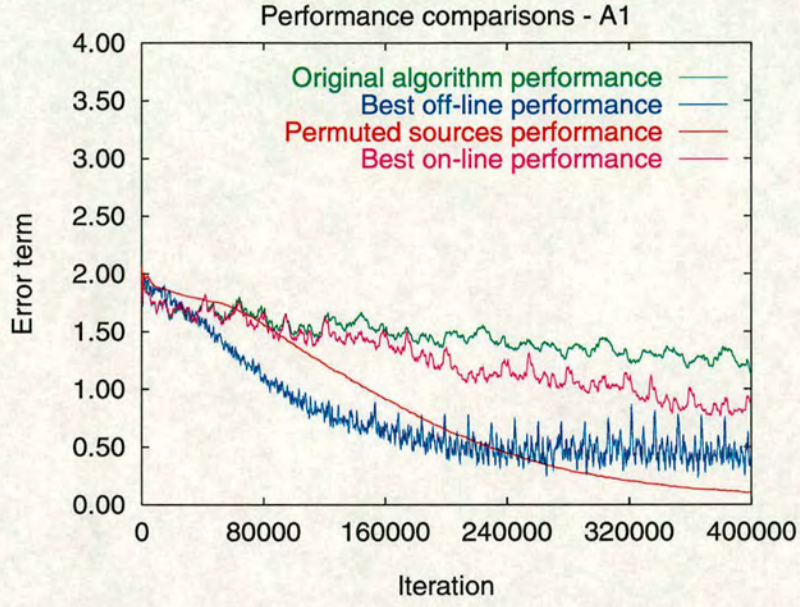


(a) Performances for **A1** mixtures

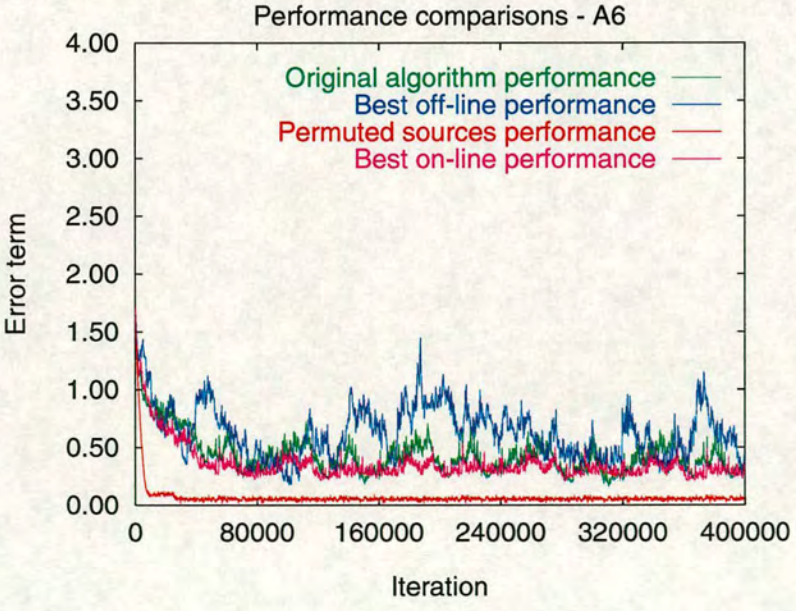


(b) Performances for **A6** mixtures

Figure 5.4: *Best and worst performances*



(a) Performances for **A1** mixtures



(b) Performances for **A6** mixtures

Figure 5.5: *Performance comparisons*

1.00 and 0.50 s, or thresholds of 5.00% and 10.00%, was the on-line version of the algorithm worse than the original algorithm without the pre-processing. Convergence was slower than for the off-line pre-processed cases because the new system required the network to acquire the sources before it could fully take effect. This was to be expected, since the suspension of updates during periods of silence further delays the convergence process. Although techniques similar to those considered in Section 4.4.1 are being used, the differences in the results occur because the off-line sources have had all of the periods of silence removed before being mixed. Therefore, once presented to the network, every set of input samples can be used. The same is not true in this on-line version, since some of the sets presented will fail the non-stationarity assessment. Consequently, after a fixed number of input samples, the on-line network will have updated its weights less often than the network in the off-line experiments.

Another characteristic in the performance graphs of the on-line assessment system (see Figure 5.5) is that there is substantially less variation in the degree of separation of the outputs. This is of particular interest for the use of such techniques in a real-world application, as it means that the quality of the separated signals remains more constant than before. The level of noise seen in the averaged graphs was also reduced.

5.4 Additional performance-improving strategies

Having successfully incorporated the non-stationarity reduction techniques into an on-line system, investigations were initiated into improvements that could further increase the convergence rate for the non-stationary signal separation. The on-line approach suspends updates whenever one or more of the estimated source signals are determined to have undergone a significant change of variance, and as the network does not update its weight set during these periods, the approximation of the separating matrix is not improving. Hence convergence is slower than in the off-line separation case. Methods to improve this situation include enabling the network to update its weight set more often — either by modifying the inputs, providing alternative updates, or substituting data exhibiting more stationary characteristics upon which the updates can be calculated — or by ensuring that the updates that it does make are optimal.

Since the infomax-based separation system is based around a gradient ascent algorithm, many of the standard ANN algorithms and techniques developed for other purposes could be applied

here. Most of the algorithms attempt to accelerate or optimise the learning by making use of second order or conjugate-gradient based approaches to find the steepest gradient and move the solution in this direction. Whilst these and other similar approaches can offer greatly improved convergence rates, they are often very complex or computationally expensive to implement. In this study, several less complex approaches have been investigated. These techniques are based directly around the observed signal data.

The relative- or natural gradient approach is considered separately in Section 5.11.1.

5.5 Average update

Omitting an update when the outputs fail the stationarity assessment is equivalent to adding an all zero update to the weight set. To overcome this problem, the strategy of supplying an alternative update was considered. This replacement was designed not to push the weight set into a large change of value, otherwise it would be less beneficial than taking no action. The initial consideration was to re-use the previous update. This will not always be advantageous, since if it happened to be leading away from the desired solution, then repeating this step will move the weight set still further away. If, however, it was in a beneficial direction, re-using it will improve the convergence rate by a factor proportional to the relative number of samples that pass the stationarity assessment. Consequently, the advantages of this method depend to a large extent on the accuracy and validity of the stationarity assessment test. If it does not reliably identify the onset of source signal non-stationarity, or if the updates calculated immediately prior to this are not improving the estimate of the solution, then the values stored and used as replacements at each iteration during the period of output silence may result in worse convergence performance.

A second and alternative extension to this approach is to base the replacement on the average of the last n updates, where n is a parameter that is determined by experimentation, later in this section. The ideal value of n will be related to the sample rate of the signals in question and any inherent stationarity-related characteristics, such as the 20–25 ms period for speech signals.

Using this approach, the overall trend of the learning should lead towards the desired solution — if the average is taken over a sufficiently long set of the updates that exhibit this trend, it should lead in the general direction of the solution. This approach is better than using the last update alone, provided the average does characterise this trend, since all of the replacement

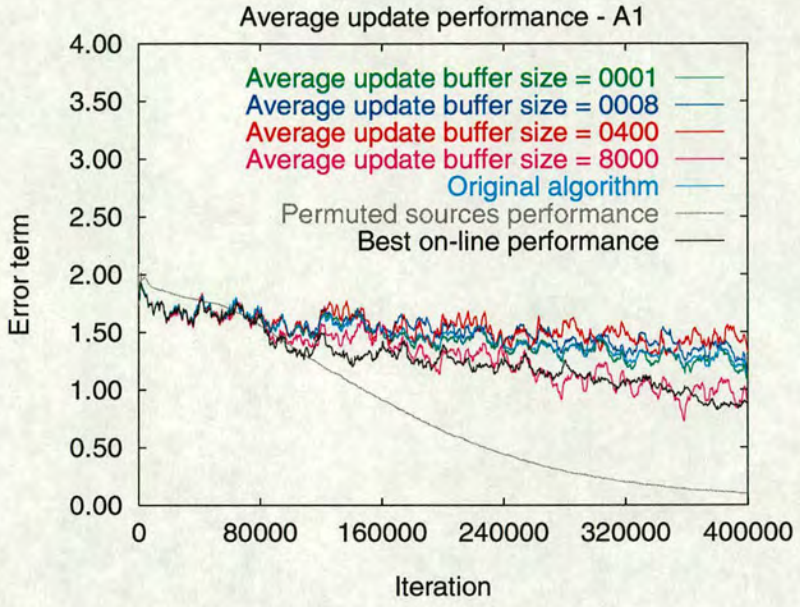
updates should then improve the convergence towards the desired solution. (For a more general framework than the one used here, if the mixing process is not static, averaging would have to be carried out over periods shorter than that over which the mixing changes.)

5.5.1 Experimentation and results

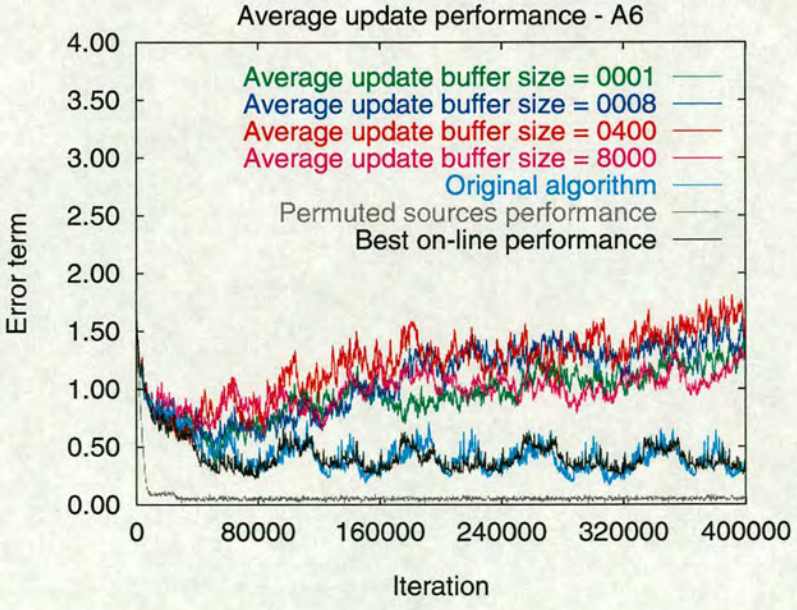
All of the experiments performed here used the same setup as in Section 5.2.2, but extended to make use of the new averaging technique. Each time an input sample satisfied the assessment criterion, the updates generated based on it were inserted into a buffer of the specified length n, and the average values of each component of the update calculated. Then, whenever an input sample failed the assessment, these average values were used as the network updates. The silence assessment parameters used were set to a duration of 0.005 s and a threshold level of 1.00%.

Typical results for a range of buffer sizes, including $n = 1$ which is equivalent to the *last update* scenario described above, as well as $n = 8$, $n = 400$ and $n = 8000$ are given in Figure 5.6. For each of the mixes, contrary to the hypothesised behaviour, the effect of the averaging and use of just the last update appeared to degrade the system's separation performance. This effect was less pronounced for the least well-conditioned mixing set (Figure 5.6(a)) due to the greater difficulty experienced by the network in trying to separate this mixture generally — hence the impact of any additional techniques is lessened. The performance loss appeared to occur with a fixed pattern related to the buffer length over which the updates were averaged. For both of the sets of graphs shown, it can be seen that the longest buffer size (in these cases, 8000 samples, equivalent to 1.000 s worth of data) gave the best performance, although this was not significantly different at the 5% level. In the best case, that for mixture **A1**, this was as good as the optimal on-line performance, but worse for each of the other cases examined. The worst performance occurred with a buffer length of 40 samples (0.005 s) and is worse than that of the original unmodified algorithm. The drop in performance may have been due to the degree of non-stationarity present in the signals over these durations, which affected the combination of updates over the different buffer lengths, as follows :

Short buffer lengths (< 20 **ms**) The buffer contains only localised data, and the average trend based on this may not lead towards the desired solution from the current location. The average may however provide some assistance in converging. This is dependent on the



(a) Performances for **A1** mixtures



(b) Performances for **A6** mixtures

Figure 5.6: *Average update performance comparisons*

length of the period of silence, and on how far the network's state is moved from a position at which the updates are beneficial before it can calculate an accurate update again.

Intermediate buffer lengths (> 20 **ms,** < 0.1 **s)** The buffer may now contain updates from sections of the inputs on either side of a typical period of silence, each potentially with a different trajectory to the solution. However, as there may be insufficient information to calculate a useful average, the combination of these updates may not directly follow any of these trajectories.

Long buffer lengths (> 0.1 **s)** The buffer should now contain sufficient updates for their average to lead towards the desired solution. This average may be beneficial for short periods of time immediately following the start of an identified period of silence, but will lose its usefulness the longer the period of silence continues. After a long period, it may become detrimental to the convergence, as the network could be far from its original location.

It would be necessary to determine for how long the update remains useful before such a technique could be safely used in a general case. A time decaying weighting of the average may suffice for this purpose, but was not tested in this study.

5.6 Buffered inputs

A buffered input strategy provides alternative data from which to construct the updates, rather than providing an alternative update for the weight set. The only requirement on this data is that it should be provide a more stationary continuation of the data immediately prior to the period of silence, such that the section under assessment now satisfies the stationarity assessment. This could be achieved in a number of ways :

- modelling the estimated source and generating typical data in the desired range
- scaling the current input to a value closer to the desired range
- using a previous data value that better fits the range

The last alternative was selected for investigation as the least complex method which required only buffering suitable data, and then selecting one or more elements of that data. The key

points of this method are the identification of suitable data to store, the size of the storage buffer and the method by which the samples are selected.

Identification of suitable data was not difficult, since the non-stationarity assessment methods could again be used. However, in this technique, data assessed as stationary was continually stored in the buffer, to be used the next time a section of the signal failed — in this way, the replacement data should show very little difference in characteristics from the samples immediately prior to those being replaced.

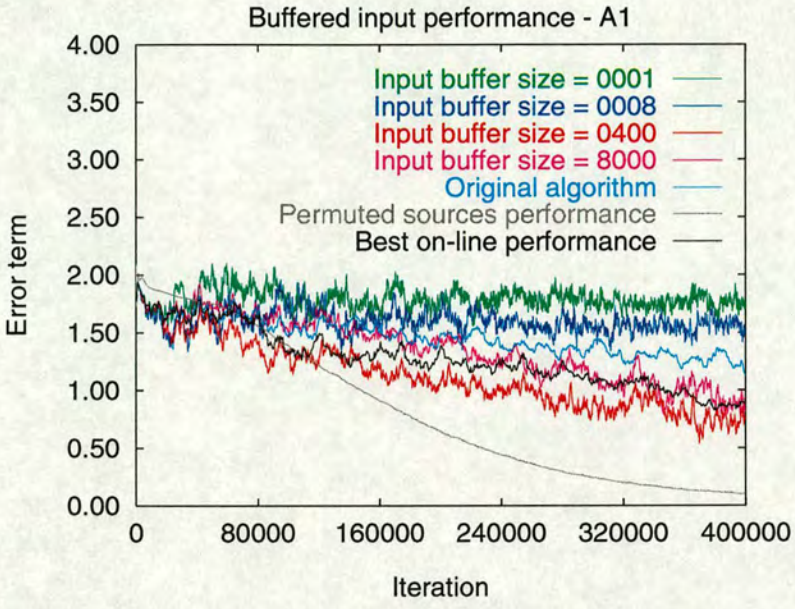
The storage buffer had to be long enough to span the likely period of non-stationarity, since although it would be possible to pick elements from the buffer multiple times, a better spread of values would give a more accurate representation of the signal, leading to better convergence to the desired solution. The experiments detailed below were designed to indicate an appropriate order of magnitude for the buffer size, n.

Selection of the picking strategy should attempt to minimise any degree of non-stationarity remaining in the replacement data — therefore a selection of (pseudo-)random samples should give the best performance. However, a simple selection strategy was used here, so that the results of the buffering were not affected by the permutation.

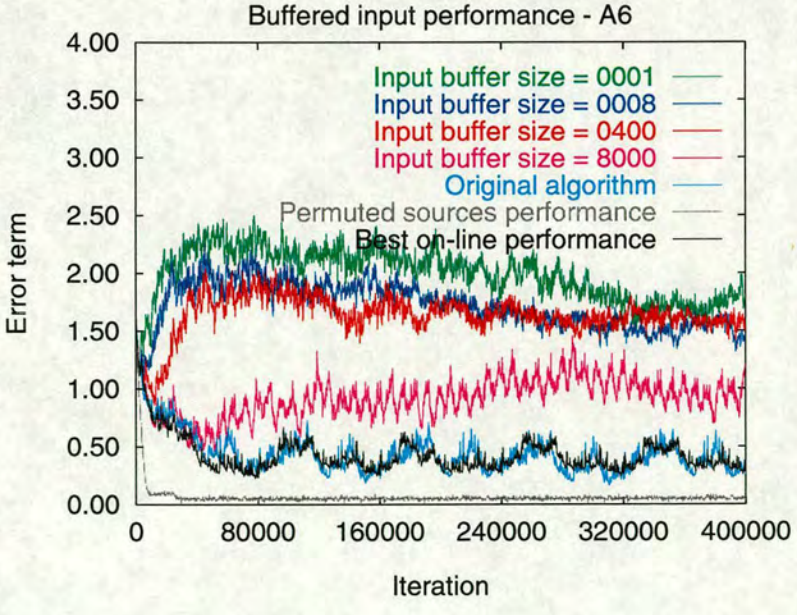
5.6.1 Experimentation and results

The simulation framework was modified to incorporate the new buffered input method created. This required the addition of extra buffers to hold the input signals whenever the output assessment showed the outputs to be stationary. In this design, each time the buffer was updated, the full length of the buffer was over-written with the stored data, ending at the current sample. Whenever the outputs failed the silence assessment test, stored input data were read from the buffer, starting at a point determined by the current time index. This procedure was followed so that the same sections of data were not used repeatedly as the replacements. Successive buffered data samples were used until the output assessment again satisfied the stationarity assessment criterion.

Results for a range of buffer sizes ($n = 1$, $n = 8$, $n = 400$, $n = 8000$), run with the same assessment parameters as in Section 5.5.1, are graphed in Figure 5.7 along with the best unbuffered results for comparison. The results from the graphs and the analysis show that the buffering techniques offer significant improvement at longer buffer sizes, for the more



(a) Performances for **A1** mixtures



(b) Performances for **A6** mixtures

Figure 5.7: *Buffered inputs performance comparisons*

difficult mixtures. Again, this is due to the rate at which the algorithm adapts to the changes in the mixture, limiting the effect of the improvements under consideration. For the more well-conditioned mixtures, the short buffer size simulations fared badly, the likely reason being that the methods used to select the sections of input to store cause the buffering of those parts of the signal that have high energy. These may have been sections that were abnormally loud, and not particularly characteristic of the mean or variance of the signal — hence there may still be a considerable amount of non-stationarity as the system switches between the current and buffered inputs. The intermediate and long-sized buffers gave the better performances for the less well-conditioned mixing matrices, showing significantly improved performance over the unadorned on-line silence removal system. The longer buffers should contain sections of the inputs of more constant, higher amplitude, due to the selection process, and thus the substitution of these samples during periods of estimated silence should provide a more consistent trend in signal, leading to better convergence.

This difference in performance between mixtures limits the usefulness of the technique for the same reason expressed previously — that no assessment of the conditioning of the mixture can be made in a truly blind situation.

5.7 Variable learning rate

Experiments carried out on the variable learning rate experiments used the non-stationarity assessment in a different way. Rather than suspending the updates, or replacing them with alternative data or updates, the degree of non-stationarity was used to control the learning rate parameter of the network's update rule. Combinations of the assessment values from each of the estimated sources can be used in different ways to achieve different effects, but the most straightforward approach is a simple summation of the values of each of the signals. The ratio of this sum to a set level (such as the maximum possible level) is used to determine the step size, so that the more non-stationary the signals, the smaller the step size used. The usefulness of this technique depends on the assessment method producing a measure than can be compared against a reference value, which is satisfied by the non-stationarity assessment criterion previously used. By this methodology, the effect of non-stationary data that would otherwise lead to the large changes in the values of the weights can be constrained, whereas the more stationary data is not penalised. The convergence could even be accelerated during such periods, depending on the method of combination of values, and the comparisons performed.

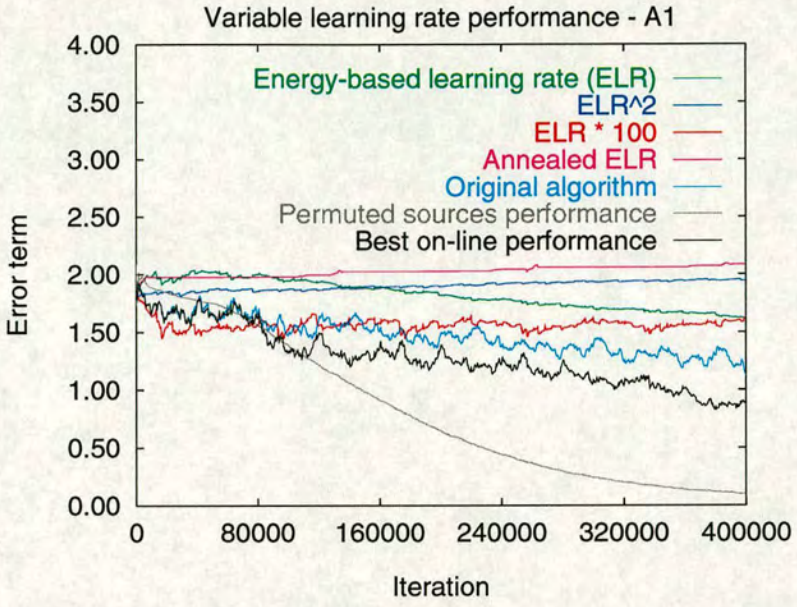
In some ways, this technique is similar to the effects of simulated annealing, often used for refining iteratively calculated solutions in artificial neural networks [14, 18]. The annealing involves reducing the learning rate of the update algorithm either after a certain period of time, or after a desired level of performance has been reached. This reduction may allow the system to settle more closely to an extrema around which it may be oscillating. It may be possible to combine these two methods to further improve the performance of the system, providing an appropriate stage at which to anneal could be determined. Other studies using a variable learning rate includes [121].

5.7.1 Experimentation and results

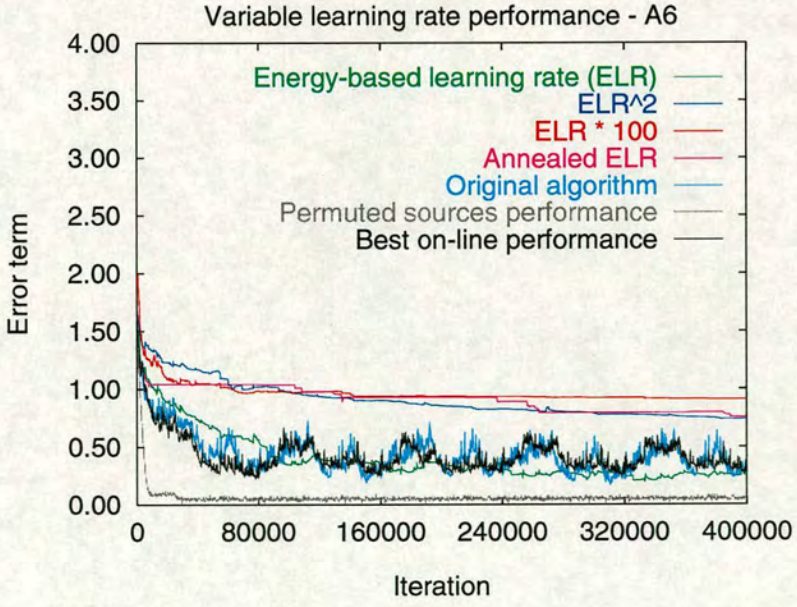
These experiments again used the same basic simulation framework as described in Section 5.2.2. The method used for determining the network learning rate was :

1. Determine the instantaneous energy of each of the outputs : $\mathbf{E} = \mathbf{Y(t)}^2$
2. Calculate a value for half of the maximum possible energy of each output, based on the maximum value seen so far : $\mathbf{E_{th}} = \frac{1}{2}\mathbf{Y}^2_{\mathbf{max\_seen}}$
   (Use of maximum values is required to take account of the indeterminate scaling. Halving them allows a ratio greater than $1:1$ to be produced, thus accelerating learning.)
3. Divide the former by the latter to give a current ratio of the output energy to the set level for that channel (this will change as the maximum value seen so far changes) : $\mathbf{R} = \frac{\mathbf{E}}{\mathbf{E_{th}}}$
4. Sum the ratios generated, and divide by the number of outputs (and optionally scale or post-process the resulting value by a function $f()$ to further improve the eventual rate modification factor — see later comments) : $scale = f\left(E\left\{\mathbf{R}\right\}\right)$
5. Multiply the basic learning rate (LR) by this modification factor, and use this final value as the network's learning rate : $ELR = LR * scale$

Several different tests were run, with the different scaling or post-processing techniques described above applied at the penultimate stage of the above procedure. The results of all of these tests for the **A1** and **A6** matrices are given in Figure 5.8, along with the plain on-line silence removal results and the original permuted source results, for comparison. The learning rate rules graphed are :



(a) Performances for **A1** mixtures



(b) Performances for **A6** mixtures

Figure 5.8: *Variable learning rates performance comparisons*

ELR The energy-based learning rate, described above.

ELR ^ 2 As before, but the scaling factor is squared before the final multiplication.

ELR * 100 An amplified version of the ELR to investigate the effect of order of magnitude.

Annealed ELR Annealing was applied to the ELR, halving the scaling factor every 10 seconds. (This left a reasonable range of learning rate values over the duration of the simulation — for a continuous system, a more suitable annealing strategy should be investigated.)

Graphs (Figures 5.8(a) and 5.8(b)) show that none of these scaling or post-processing techniques, such as the squaring or annealing, further assisted the separation performance beyond that of the original ELR.

Figure 5.8(a) has to be interpreted carefully. Although the graph indicates relatively poor separation performance for the variable learning rate methods compared to the other approaches, the statistical analysis of the end points of the extended simulations show significant improvement over original algorithm performance. For each of the other mixtures considered, the energy-based learning rate (ELR) method's performance was closer to that of **A6** shown in Figure 5.8(b).

5.8 On-line permutation

To provide a comparison of the effectiveness of the above techniques, an alternative method for improving the performance of the information-maximisation algorithm on non-stationary sources was devised. The method, on-line permutation, is based on Bell & Sejnowski's original work.

In their experiments, Bell & Sejnowski [10] pre-processed the source signals by permuting them to ensure stationarity prior to mixing. This is not possible in a truly blind situation, since nothing is known about the source data, nor their characteristics — only the input signal can be manipulated. Permutation of these inputs will not lead to stationarity of original sources, but may help reduce the non-stationary effects to some extent. Since the entire input signals cannot be known in advance, the best that can be achieved is a partial permutation, carried out over a 'sliding window' of buffered sections of the input signals.

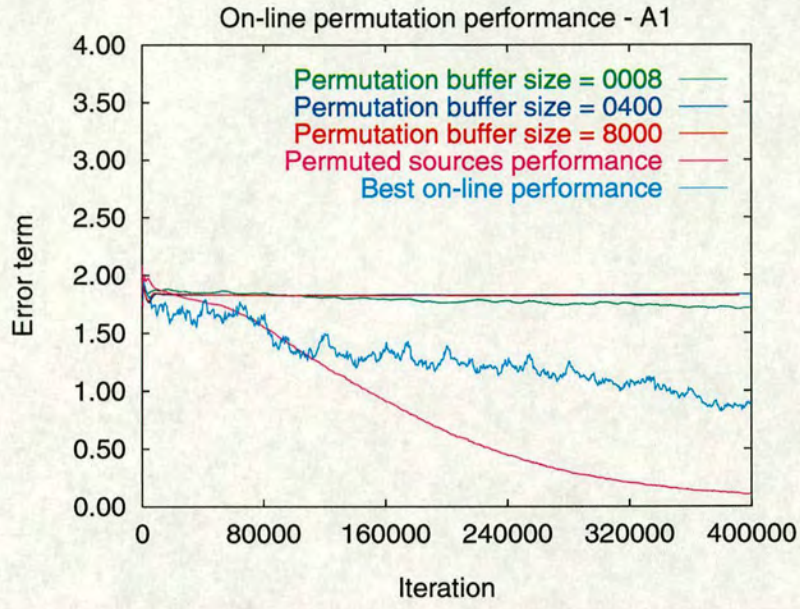
Questions to be addressed at this stage concerned the size of the buffer or window that gives the greatest improvement, and the most efficient method of generating a suitable pseudo-random sequence for selecting elements from this buffer. In this study, the issue of buffer size was found by experimentation and a scaled univariate pseudo-random number generator was used to generate the permutation order. Once the buffer had initially been filled, elements were drawn from it according to this pseudo-random order, and the new input data were inserted into the newly vacated slots.

5.8.1 Experimentation and results

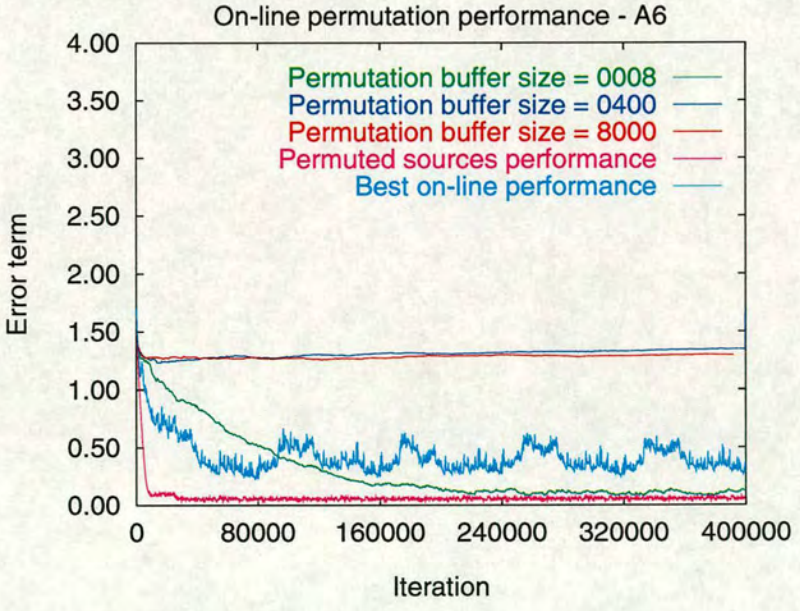
Tests were run to determine the optimal buffer size over which to carry out the on-line permutation of the input signals. Buffer sizes under consideration ran from 8 to 8000 samples (0.001 to 1.000 s), covering the range of durations for consideration of the stationarity of the signals. The experimental setup was modified to read the incoming data into an appropriately sized buffer and then randomly pick corresponding sets of input data, one from each channel, to present to the network.

Illustrative results from the middle and extremes of the range of values tried are graphed in Figure 5.9 for the least and most well-conditioned matrices. The results from the graphs showed that the on-line permutation did not perform well at any buffer sizes for the more difficult mixtures of the **A1** mixing matrix, but performed considerably better on the easier mixtures of the **A6** matrix, for small buffer sizes. The algorithm fares better over shorter buffer sizes because the short buffer size corresponds to a period below the speech stationarity threshold, which means that the source signals, and consequently their linear mixtures, can be considered stationary over this period.

Under these conditions, the level of separation achieved was close to that attained when the sources are permuted prior to mixing, although the on-line permutation system takes longer to converge. It is, however, better than the results from the on-line silence removal tests in Section 5.2.5. For the more difficult mixtures, the performance was worse than that of the unmodified infomax algorithm on the unprocessed source mixtures. The poor performance is due to the fact that the permutation makes it more difficult for the algorithm to converge under these circumstances. Convergence would normally be slower for these more difficult mixtures, and now in addition the weights are continually fluctuating with little chance to start to converge (even briefly) before the mixture changes again.



(a) Performances for **A1** mixtures



(b) Performances for **A6** mixtures

Figure 5.9: *On-line permutation performances*

For the other mixing matrices with determinants between those of the two illustrated above, the technique performed reasonably well. However, due to the poor performance of this technique for the least well-conditioned case, and the impossibility of predicting the degree of mixing in a truly blind situation, this approach cannot be recommended unless some way of overcoming, or compensating for this situation can be found.

5.9 Comparison of alternative update strategies

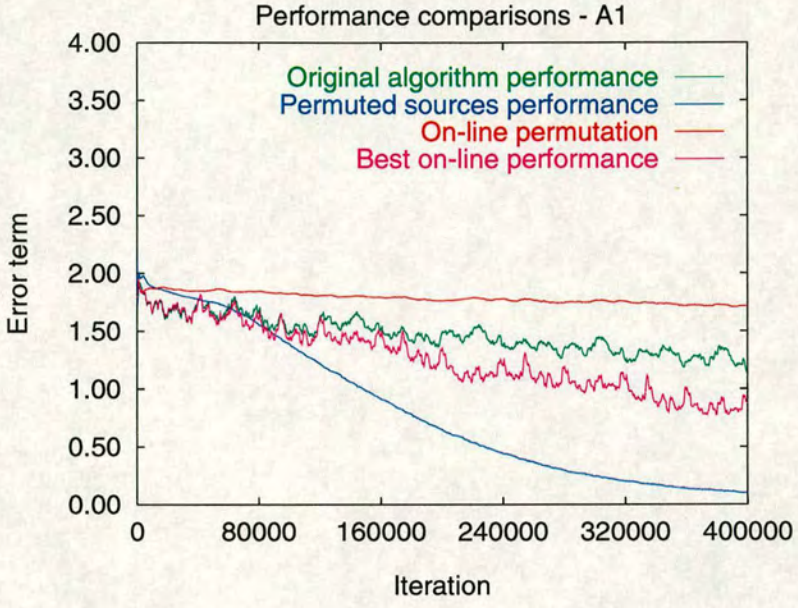
To complete the section on alternative update strategies, a simple comparison of the best performances achieved by each method is presented. The graphs in Figures 5.10 and 5.11 show the performance of the various techniques for the **A1** and **A6** mixing matrices. The graphs in Figures 5.10(a) and 5.11(a) show the separation levels achieved by the original algorithm on inputs created from both the unprocessed and permuted sources, the results of the on-line permutation, and the best performance of the modified infomax algorithm on the inputs from the unprocessed sources. The graphs in Figures 5.10(b) and 5.11(b) repeat this last result for comparison, with the levels attained by the various alternative update strategies.

It can be seen that none of the methods tested out-performed the case where the inputs were created from the permuted source signals. This result could be expected, due to the superior non-stationarity elimination. From the graphs, the on-line permutation approach and the variable learning rate method appeared to offer similar levels of performance for the most well-conditioned mixing case, albeit taking longer to converge, but performance on the least well-conditioned system appeared slightly worse than that of the original infomax algorithm.

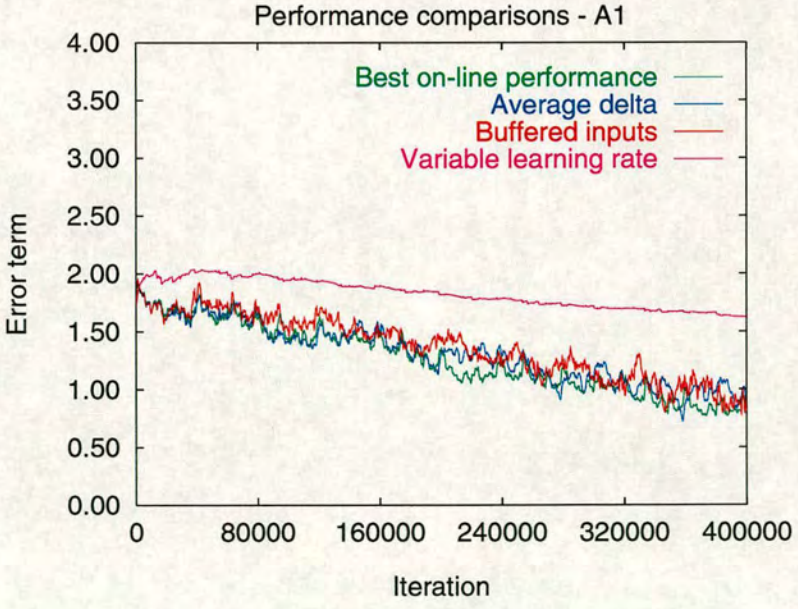
Use of the average delta and buffered input strategies proved not to be generally helpful in these experiments, over the range of buffer sizes investigated, when considered for all mixing matrices. The modified infomax algorithm that incorporated the on-line non-stationarity reduction did not a significant improvement in separation performance.

5.10 Performance on multiple inputs

To examine whether the performance improvements proposed in the earlier sections could be extended to work with larger network sizes, tests were run using a five input network. The five source signals used were those described in Section 4.6.4, *i.e.* the two supplied by BT and used

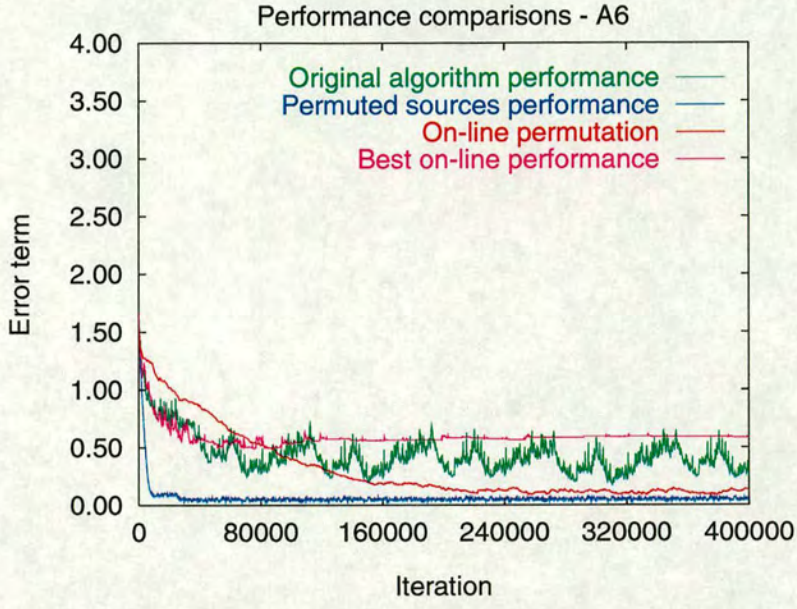


(a) Performances for **A1** mixtures

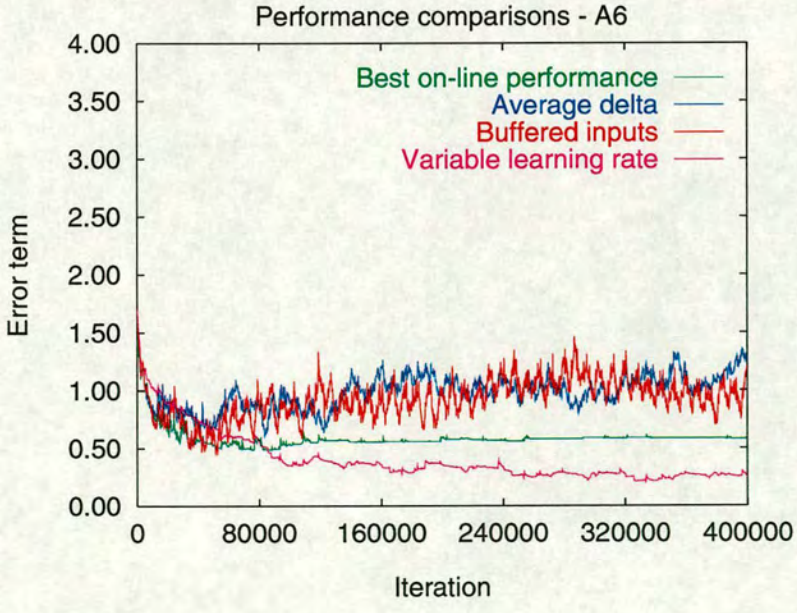


(b) Performances for **A1** mixtures

Figure 5.10: *Comparison of alternative update strategies' performances for mix* **A1**



(a) Performances for **A6** mixtures



(b) Performances for **A6** mixtures

Figure 5.11: *Comparison of alternative update strategies' performances for mix* **A6**

in all of the other experiments, plus the three recorded independently. The signals were mixed using the matrices given in Table 5.2, selected from a set of appropriately-sized matrices of univariate random numbers in the range $[-1.00000 : 1.00000]$, with their determinants given in Table 5.3. The matrices chosen were considered to give a reasonable spread of conditioning

$$\mathbf{A7} = \begin{bmatrix} -0.83847 & -0.92972 & 0.12772 & 0.24506 & -0.66833 \\ -0.66257 & -0.83975 & -0.29840 & 0.50092 & 0.81378 \\ 0.66394 & 0.34174 & -0.47839 & 0.02103 & 0.99484 \\ -0.35641 & 0.67642 & -0.12390 & -0.12729 & -0.03943 \\ -0.00647 & 0.91022 & -0.59619 & -0.48709 & -0.99313 \end{bmatrix}$$

$$\mathbf{A8} = \begin{bmatrix} 0.12997 & 0.09665 & 0.64514 & 0.94281 & 0.49161 \\ -0.32773 & 0.38370 & -0.14166 & -0.46222 & 0.83859 \\ 0.67575 & 0.90240 & 0.66782 & -0.42816 & -0.76334 \\ 0.16278 & 0.14896 & -0.84787 & 0.83997 & -0.20956 \\ -0.52853 & -0.43832 & 0.90616 & 0.50272 & -0.55478 \end{bmatrix}$$

$$\mathbf{A9} = \begin{bmatrix} 1.0000 & 1.0000 & 1.0000 & 1.0000 & -1.0000 \\ 1.0000 & 1.0000 & 1.0000 & -1.0000 & 1.0000 \\ 1.0000 & 1.0000 & -1.0000 & 1.0000 & 1.0000 \\ 1.0000 & -1.0000 & 1.0000 & 1.0000 & 1.0000 \\ -1.0000 & 1.0000 & 1.0000 & 1.0000 & 1.0000 \end{bmatrix}$$

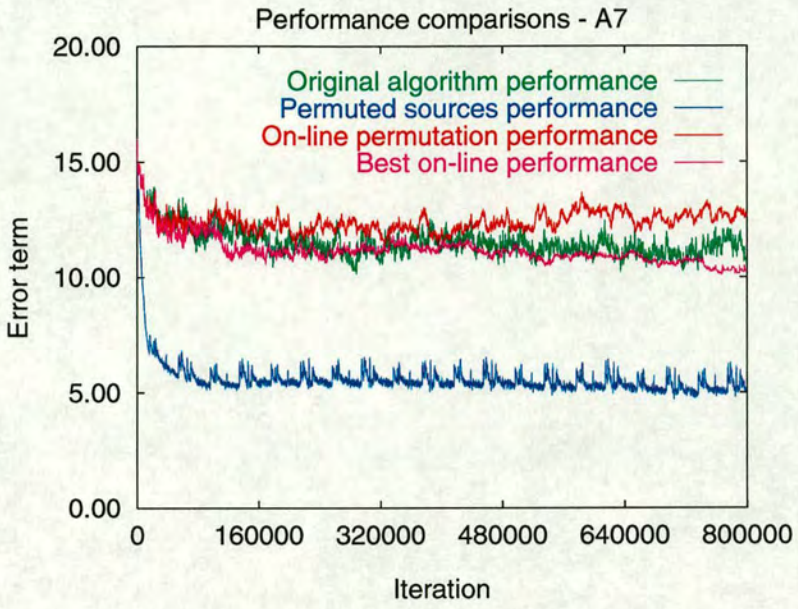
Table 5.2: *The 5 by 5 mixing matrices*

$\det(\mathbf{A7}) = -0.02789$	$\det(\mathbf{A8}) = 1.63218$	$\det(\mathbf{A9}) = 48.00000$

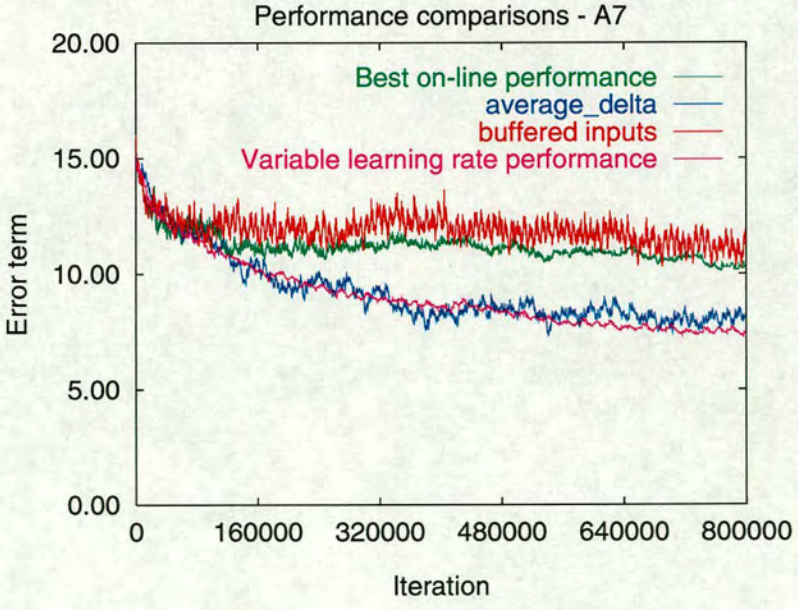
Table 5.3: *Determinants of the 5 by 5 mixing matrices*

over which to assess the separating performance, for these networks. As before, random initial values for the weight matrices and bias vectors were generated.

Due to the increased size of the networks and the associated complexity of solving the separation problems, simulations were this time run for twice as long as the previous cases — 100 seconds. As well as examining the performance of the on-line silence removal algorithm, the average delta, buffered input, variable learning rate and on-line permutation systems were also tested. These were compared to the performance of the original infomax algorithm on the

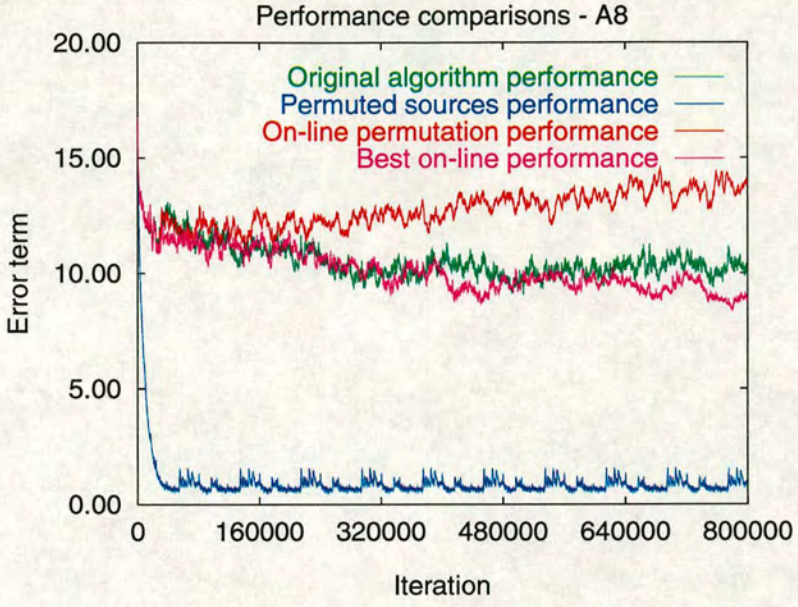


(a) Original, permuted, on-line permutation and on-line silence removal performances for **A7** mixtures

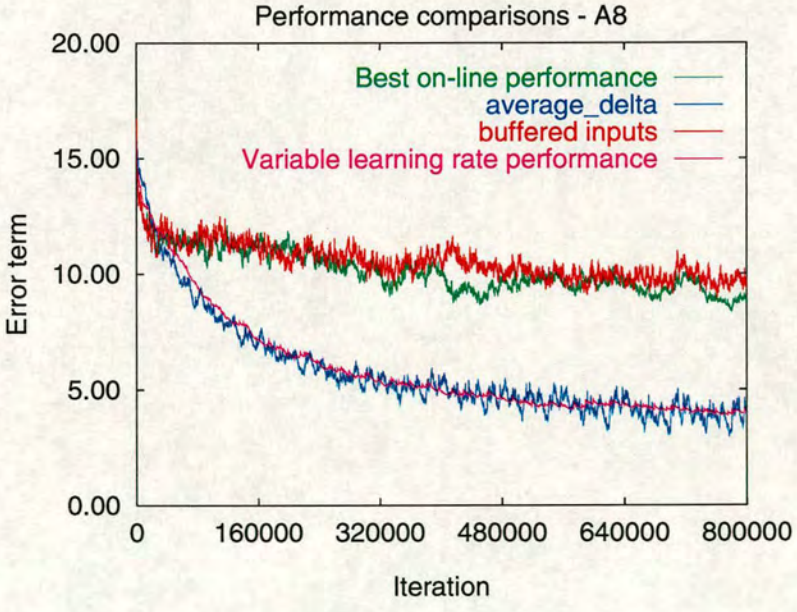


(b) Average delta, buffered inputs and variable learning rate performances for **A7** mixtures

Figure 5.12: *5 input separation results for mix* **A7**

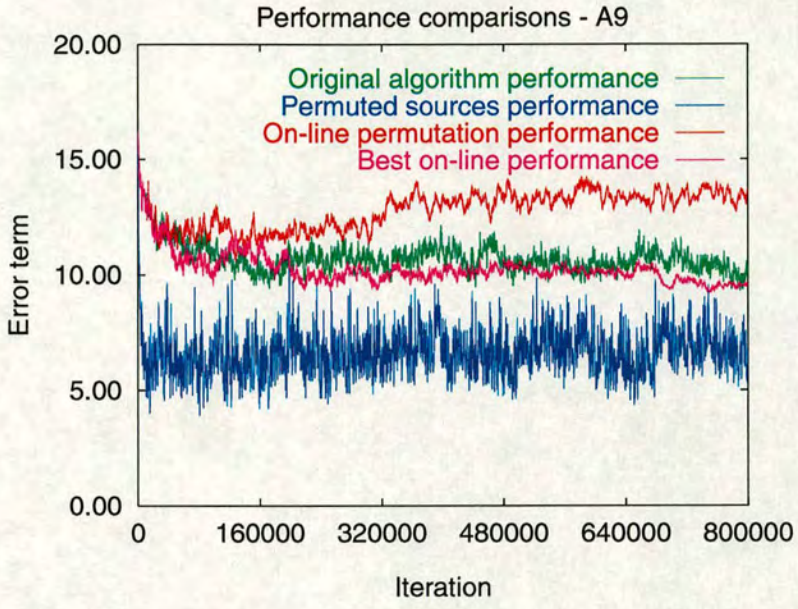


(a) Original, permuted, on-line permutation and on-line silence removal for **A8** mixtures

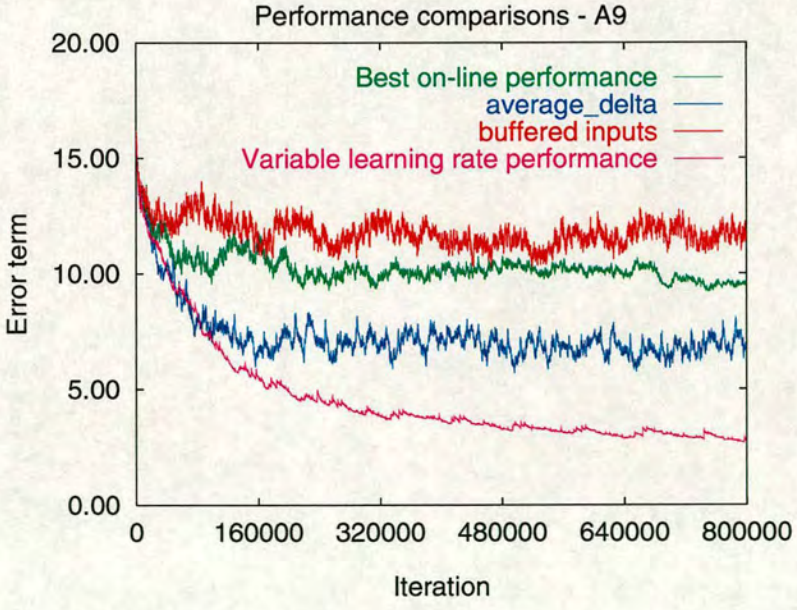


(b) Average delta, buffered inputs and variable learning rate performances for **A8** mixtures

Figure 5.13: *5 input separation results for mix* **A8**



(a) Original, permuted, on-line permutation and on-line silence removal for **A9** mixtures



(b) Average delta, buffered inputs and variable learning rate performances for **A9** mixtures

Figure 5.14: *5 input separation results for mix* **A9**

unprocessed sources, and on the permuted sources. Results for all of these experiments are given in Figures 5.12, 5.13 and 5.14.

The graphs of the results of the basic comparisons (Figures 5.12(a), 5.13(a) and 5.14(a)) show that the separation performance of the permuted sources is far better than that of the unprocessed signals, and this is confirmed by the statistics. In the last case (for mixing matrix **A9**) the convergence trajectory is very noisy. This is most probably due to the fact that the mixing matrix **A9** is very well conditioned, and therefore the algorithm is capable of tracking the changing characteristics of the sources rapidly — hence it leaps between trajectories very quickly, and the resultant graph is a combination of these different trajectories, even though it appears to be a noisy representation of just one.

From observation, the on-line non-stationarity reduction approach does not appear to offer such an advantage as it did for the smaller networks, suggesting that the improvement available due to this technique is finite and does not scale linearly with the number of sources.

In the extended runs, over which the statistical analysis was carried out, the separation performance of the unmodified algorithm was better than that shown in the graphs in Figure 5.12, 5.13 and 5.14. All of the other experimental results correspond to those shown in the graphs.

When examining the performances of the alternative update strategies for the modified algorithm, the following observations can be made :

Average update This system fares far better in these tests, matching the separation level of the variable learning rate method for the first two mixtures (**A7** and **A8**), and achieving a level similar to that of the permuted sources case for the **A9** inputs.

Buffered inputs In this series of tests, this strategy again performed poorly, offering no improvement (and even some degradation in performance for the **A9** case).

Variable learning rate These results are very good, approaching the level of the permuted sources for the two more difficult mixtures (**A7** and **A8**), and far exceeding it for the most well-conditioned case (**A9**).

On-line permutation The results from the on-line permutation method for these five input tests do not show the same degree of improvement for the better-conditioned mixtures as they did in the two-input tests over the periods of the simulation. This is due to the

increased complexity of the separation problem for five signals, and is in line with the noisy nature of the off-line permutation results for the **A9** mixture.

The effectiveness of most of these techniques has been shown to extend to larger array sizes, be it beneficial, or in the case of the buffered inputs approach, detrimental.

5.11 Comparisons with other learning algorithms

A series of experiments were performed to compare the performance of these various modifications to the information-maximisation-based learning algorithm with the relative performance of other learning rules. It was also considered interesting to see if the improvements proposed for the modified infomax algorithm were could be generalised to a larger class of techniques. Four other learning rules were chosen for comparison, two of which are specifically aimed at non-stationary signals :

Natural Gradient or *relative gradient* variations of the information maximisation rule, by Amari *et al.* [58] and Cardoso & Laheld [60] respectively, enable greatly accelerated convergence compared to the original infomax algorithm (see Sections 3.4.1.1 and 3.4.3.2). This comparison investigated whether or not the non-stationarity reduction had the same effect on the convergence of this accelerated algorithm.

Hérault & Jutten's learning rule, proposed in [56] (see Section 3.4.3.1) for more details of the learning rule). This algorithm was selected as a contrasting approach since it is based on an entirely different network configuration and learning rule — one that is not entropy based, and so does not rely on the same signal criteria as the infomax algorithm.

Matsuoka, Ohya and Kawamoto's system [51] makes use of the non-stationary characteristics of the input signals in the learning rule itself. It is similar to Hérault & Jutten's architecture, but the learning rule is quite different.

Pre-filtering for non-stationary sources, due to Barros and Ohnishi [53] follows the natural gradient variation of the infomax approach, in terms of architecture and learning rule, but applies a low-pass filter to the input signals before transforming and processing them for the separation part of the system.

Whilst several other approaches exist, the above were selected to cover a representative range of methods and techniques.

For the comparison tests, the software model of the separation system described in Section 5.2.5 was modified to make use of the appropriate learning rule, and configured to make use of batch processing techniques, where appropriate. Each of the modifications and alternative update strategies described in Sections 5.5 through to 5.8 were also implemented for each algorithm. This comprised :

- 2 source signals
- 6 mixing matrices, with a range of degrees of conditioning
- 5 initial weight sets
- 6 different start points within the mixed signals
- 2 different performance metrics — Amari *et al.*'s and Barros & Ohnishi's

Initially, the results from each set of experiments were examined individually, with regard to the unmodified version of the various algorithms. The results of the respective performances, based on the graphed outputs, are given below, followed by comment on their comparative performances with each other.

5.11.1 Natural gradient

A natural gradient variant of the information-maximisation based algorithm was implemented so that the impact of the silence removal techniques could be assessed with regard to the accelerated convergence offered by this algorithm. Due to this accelerated convergence rate, initial convergence tests determined that the simulations need only be run for 240000 iterations, by which time all of the mixtures had converged to their final values (see Figure 5.15). This is considerably faster than the original infomax algorithm, for which 4800000 iterations were required to achieve convergence in some cases. It can be seen that the traces in Figure 5.15 are much cleaner than those of the infomax runs, and that for each of the mixtures, the final separation performance achieved is far better. This is again due to the improved convergence properties of this algorithm.

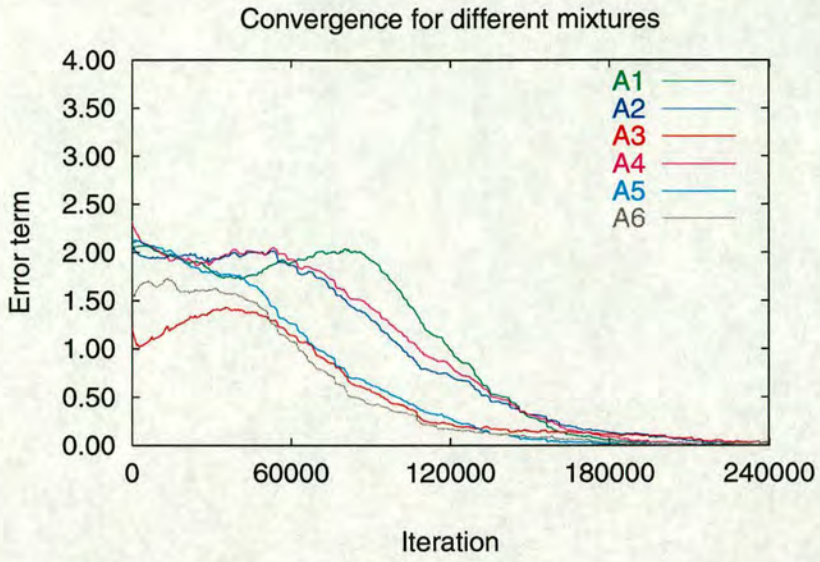
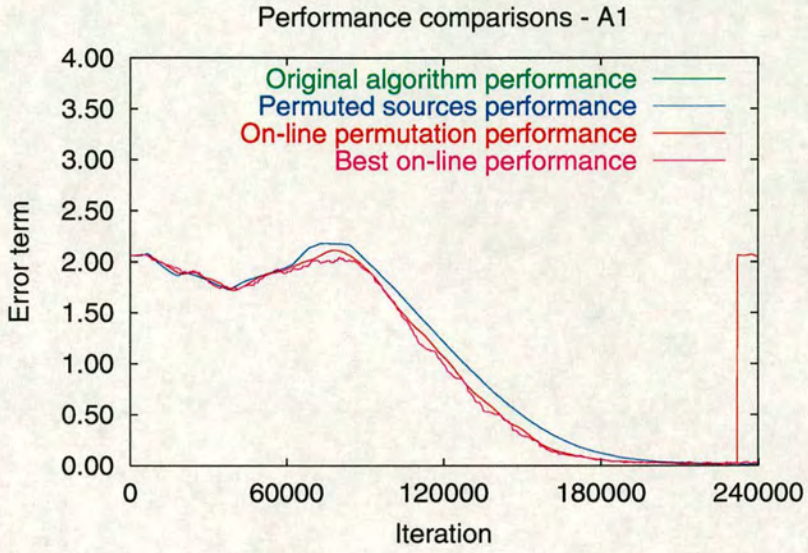


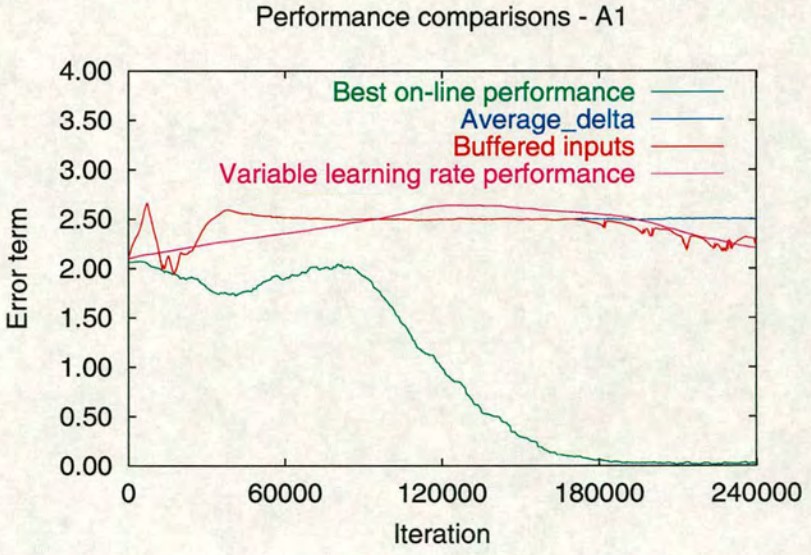
Figure 5.15: *Convergence of the Natural Gradient network*

Figure 5.16 and 5.17 show the results of the separation runs using the same modifications that had been made to the infomax algorithm for the **A1** and **A6** mixtures, respectively. Figures 5.16(a) and 5.17(a) show the separation performance for the original, unmodified algorithm run on both the mixed source signals and the mixed permuted source signals, as well as the results from the on-line permutation (for a buffer size of 8000). Also shown in these figures is the trace for the on-line silence removal results, when run with the duration and threshold parameters identified as giving the best results from all 25 combinations considered. Figures 5.16(b) and 5.17(b) show the traces for the average delta, buffered input and variable learning rate update strategies. The trace of the best on-line performance is reproduced on this graph for ease of comparison. The graphs for both of the mixtures show that there is no visible difference in the separation performance when comparing the results of the on-line silence removal modification to those of the original algorithm. The on-line permutation modification fares marginally better, but its use would not be justified in terms of the performance gain over the cost of the additional processing. Both the original algorithm and the on-line permutation traces show slightly faster convergence than that of the original algorithm on the permuted sources. Finally, all of the alternative update strategy modifications result in a considerable decrease in separation performance, and should consequently be avoided.

The reasons for the lack of difference in the traces of Figures 5.16(a) and 5.17(a) is that the original algorithm's convergence is near optimal, and hence there is little scope for any other

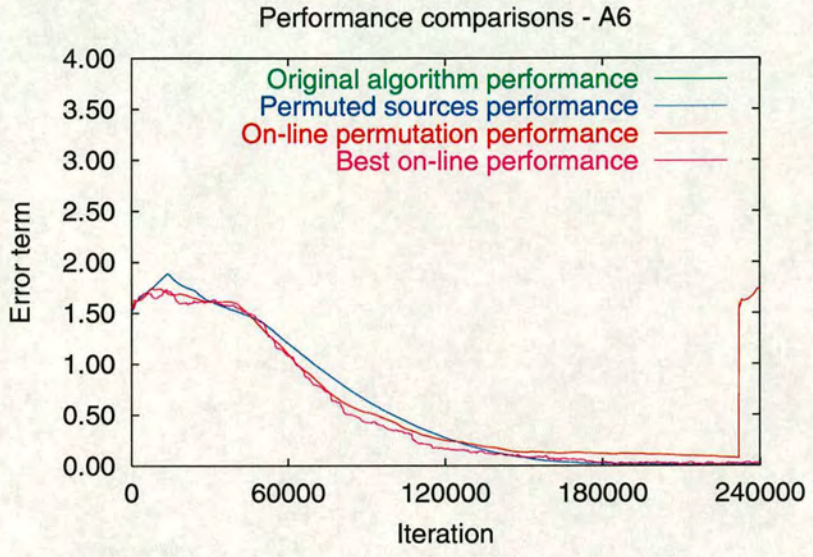


(a) Original, permuted and on-line performances for **A1** mixtures

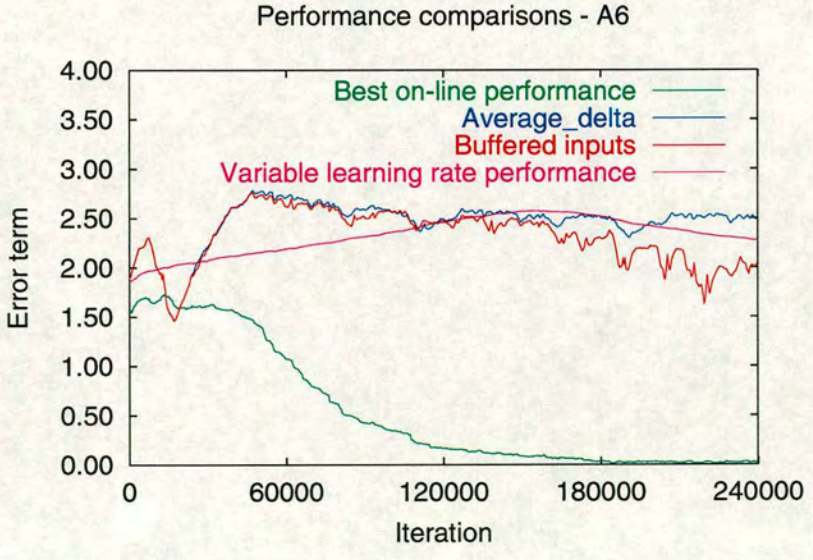


(b) Average delta, buffered inputs and variable learning rate performances for **A1** mixtures

Figure 5.16: *Separation results from the Natural Gradient network for mix* **A1**



(a) Original, permuted and on-line performances for **A6** mixtures



(b) Average delta, buffered inputs and variable learning rate performances for **A6** mixtures

Figure 5.17: *Separation results from the Natural Gradient network for mix* **A6**

technique to improve on this performance. The alternative update strategies fare so badly because they modify the algorithm's behaviour and prevent it from carrying out the updates that would otherwise lead to this good performance. Furthermore, they may present the algorithm with data that will result in it diverging from the desired convergence trajectory. Due to the fast convergence properties of the algorithm, these variations from the convergence trajectory mean that the weight set may rapidly move out the vicinity of the desired solution and in the basin of attraction of a different solution altogether.

The use of the any of the non-stationarity reduction modifications or alternative update strategies with this algorithm offers no performance improvement.

5.11.2 Hérault-Jutten

In this set of experiments, and those relating to Matsuoka, Ohya & Kawamoto's algorithm (Section 5.11.3), the learning rule takes a slightly different form to that of the infomax algorithm due to the different architecture used. (See Sections 3.4.3.1 and 3.4.3.6 for more details.)

Consequently, the performance results for the Hérault-Jutten network are generated from the matrix derived from the inverse of sum of the weight matrix and an appropriately sized identity matrix, as given by the expression $(\mathbf{I} + \mathbf{W})^{-1}$, using Amari's performance metric as before.

For three of the mixtures, **A1**, **A4** and **A5**, the results failed to attain the same levels of separation as those achieved by the infomax algorithm. This is shown in the graphs of Figure 5.18. For the remaining mixtures, **A2**, **A3** and **A6**, the results diverged.

This behaviour appears to be related to the differences in magnitude of the values in the mixing matrices. When the values are similar in magnitude, the separation performance was poor (see the mixing matrices in Section 4.6.5). In those cases, it appeared from audio assessment, that one signal was separated successfully, but that the other mixture was degraded. Within the timeframe of this study, it was not possible to pursue this further and establish the reasons for this, although it has previously been noted by Hérault & Jutten [56] that their network is capable of converging to the wrong target — *i.e.* one that is not a separating solution. The remainder of the experiments proceeded using the results recorded.

For this rule set, all of the curves were relatively noise free and settled rapidly to their final values. Those graphs that diverged did so consistently across all of the experiments using

this algorithm, subject to moderation by the extensions applied. Examination of these traces suggests that they follow a convergence path that is mirrored around the centre of the graphs shown, about the line $y = 2$. It would not be possible to accommodate such additional analysis in a real-system (*i.e.* the additional work of monitoring the outputs to look for likely cases where they have become inverted, should this occur), therefore this observation was dropped, and the traces were treated for the values they produced.

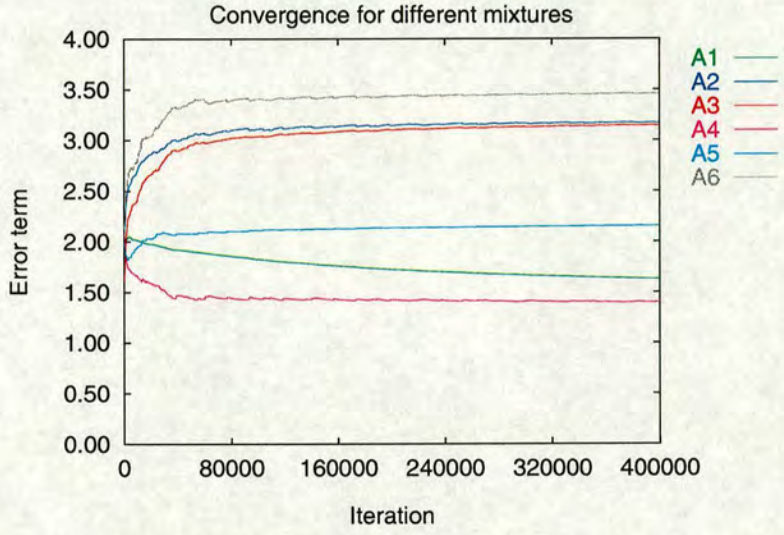
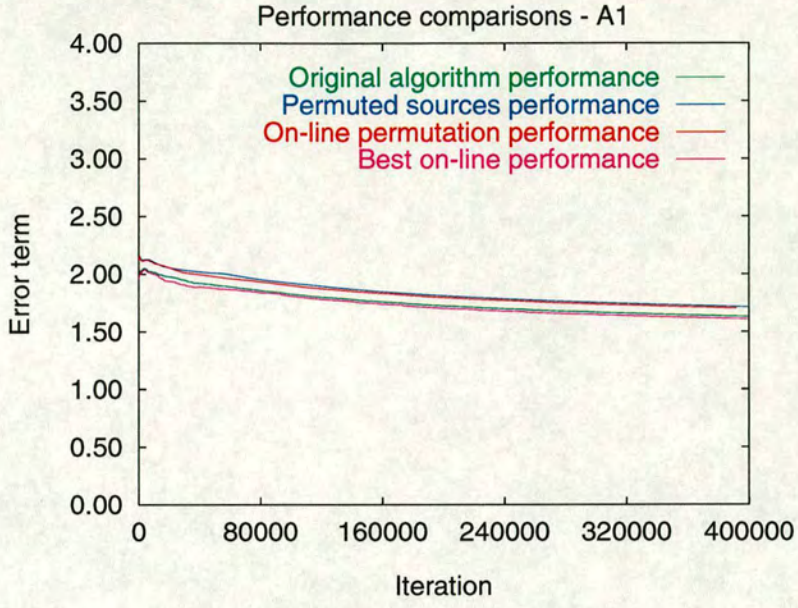


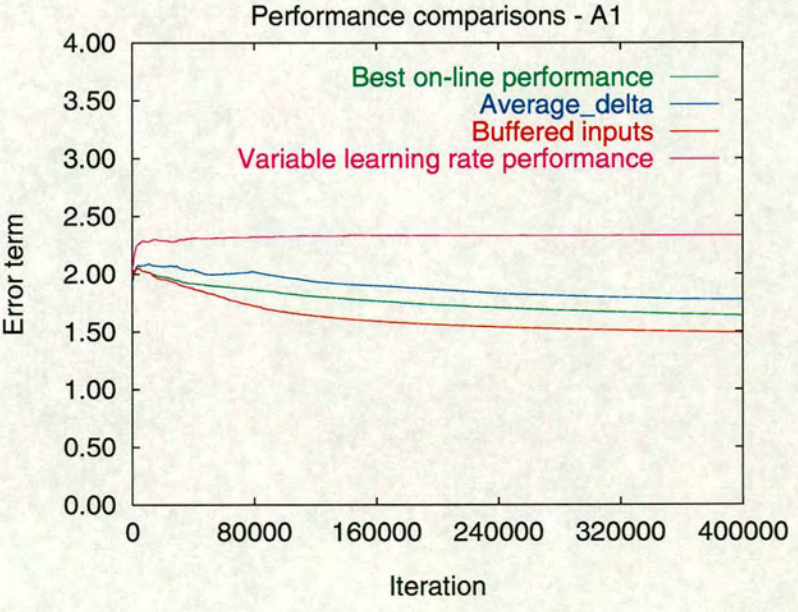
Figure 5.18: *Convergence of the Hérault-Jutten network*

The graphs in Figures 5.19 and 5.20 show comparisons of the different results from the various experiments carried out before, using the best performance for each of the two mixing matrices in the "Best on-line performance" traces, and the best overall parameters for all of the extensions. The graphs show that there was little variation in the results for the different techniques on the less well-conditioned mixtures, but that the improvement visible for the better conditioned mixtures was more marked. The on-line permutation showed a slight improvement over the original performance for both mixtures, but was not as good as that of the buffered inputs strategy. The average delta extension worked better for the more well-conditioned mixing matrix, but for the less well-conditioned matrix, it reduced the separation level achieved. In both cases, the variable learning rate approach performed particularly poorly.

The probably reason for the lack of improvement is that the non-stationarity reduction effect is small, and the effects of the extensions reflect this. The extensions and modifications can only make an impact when the on-line non-stationarity reduction causes a significant difference in

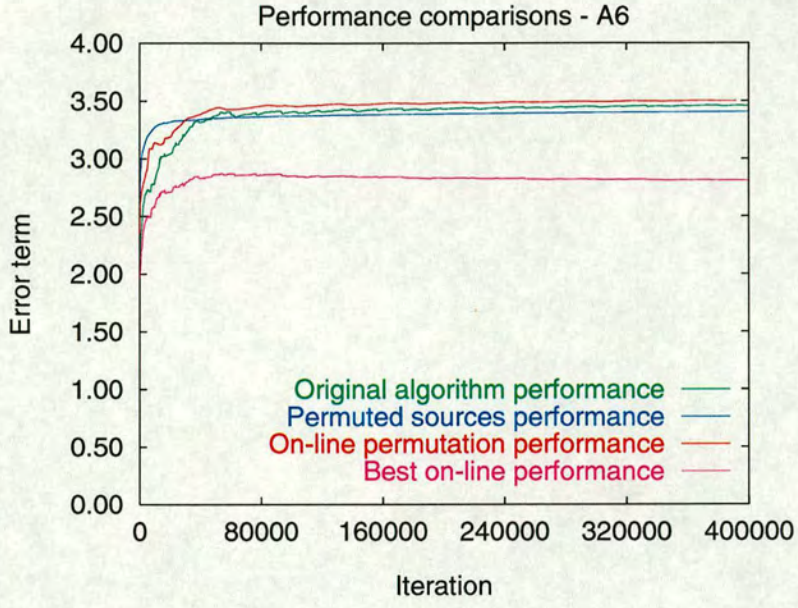


(a) Original, permuted and on-line performances for **A1** mixtures

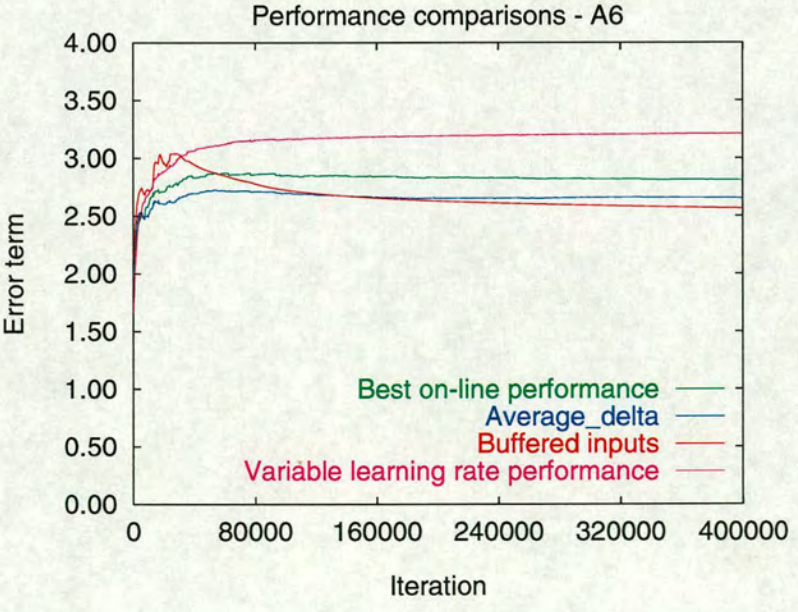


(b) Average delta, buffered inputs and variable learning rate performances for **A1** mixtures

Figure 5.19: *Separation results from the Hérault-Jutten network for mix* **A1**



(a) Original, permuted and on-line performances for **A6** mixtures



(b) Average delta, buffered inputs and variable learning rate performances for **A6** mixtures

Figure 5.20: *Separation results from the Hérault-Jutten network for mix* **A6**

performance. The reason for the poor performance of the variable learning rate extension is slightly different, since now the rapidly changing step size, related to the signal energy, merely modulates the oscillations around the convergence trajectory of the network. The amplified steps during high signal levels prevent the network settling to a solution as it should.

Thus the application of the silence removal techniques developed for the infomax algorithm do not transfer well to this architecture, resulting in only marginal gains in some of the cases, and in slight performance losses in others.

5.11.3 Matsuoka, Ohya and Kawamoto

This is the first of the algorithms considered that specifically addresses the issue of source non-stationarity. It makes use of the assumed non-stationarity in the cross-moments of the sources to drive the network weight set towards the global of a cost function derived from these values.

The network is based on a similar architecture to that of the Hérault-Jutten system, and consequently the expression from which the performance metrics are generated in this case is similar to that of the previous section, save for a transposition of the weight matrix : $\left(\mathbf{I}+\mathbf{W}^{\mathbf{T}}\right)^{-1}$. However, unlike the results from the Hérault-Jutten experiments, all of the traces from the basic algorithm did converge (see Figure 5.21), even if not in the order expected from the conditioning of the mixing matrix. As before, the graphs in Figures 5.22 and 5.23 show comparisons of the results from the experiments, using the best performance for mixtures **A1** and **A6** in the "Best on-line performance" traces, and the best overall parameters for each of the extensions. The graphs show no improvement at all to the separation performance of this algorithm, and indicate that some of the proposed modifications reduced the performance of the system. However, this result could be expected since the modifications aim to reduce the degree of non-stationarity of the signals, the very characteristic that the algorithm attempts to exploit. Consequently, it is possible that these results could be used as a confirmation of the effectiveness of the modifications and extensions. Those resulting in the worst degradation of performance should be the ones that most reduced the non-stationarity, and should correspond to those providing the greatest improvement in the infomax system.

From the tabulated results of the experiment the optimal silence removal parameters from the non-stationarity reduction stage were determined to be durations of 0.125 ms and a threshold

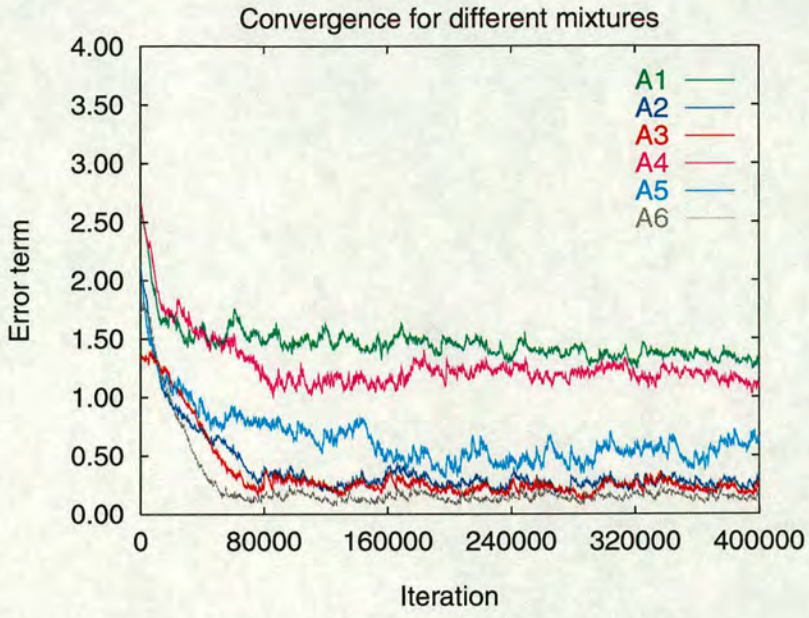


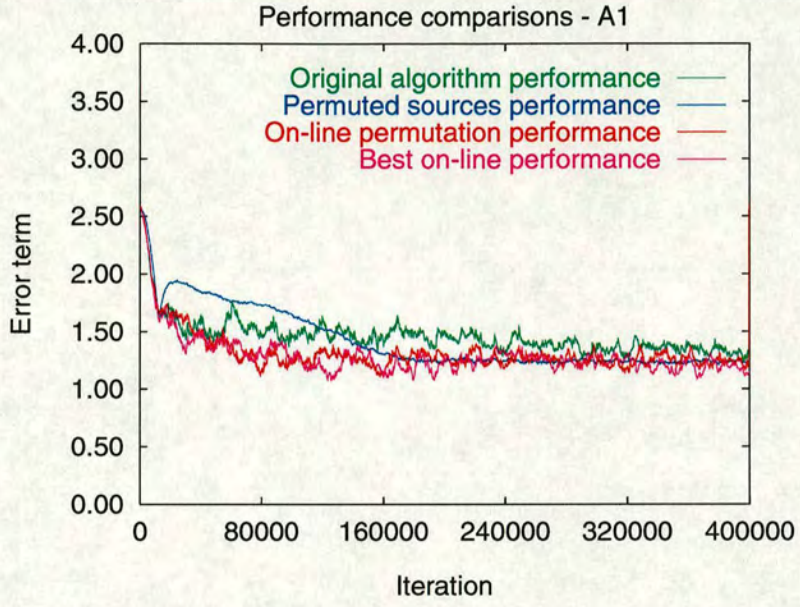
Figure 5.21: *Convergence of the Matsuoka, Kawamoto and Ohya network*

level of 0.50%. These values are quite different to those determined previously in Section 5.2.5, at the other end of the range of values considered, for both parameters — an indication that the above hypothesis is correct.

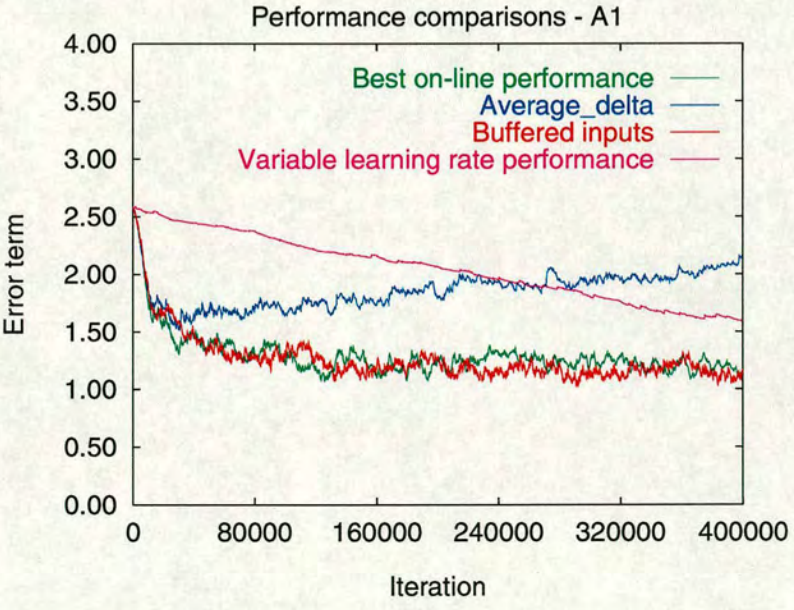
In addition, from the graphs it appears that the average delta extension had the greatest impact on performance, followed by the variable learning rate approach. The buffered inputs system was next, only noticeably different from the on-line silence removal level in the **A6** mixture. These results concur with the infomax results, but in reverse order — with the exception of the average delta method, which has had a greater effect here than it did for the other algorithms.

The reason for this latter result is that the average deltas are accumulated over a relatively long period of time — 0.1 seconds — compared to the time over which the signal is assessed by the algorithm. It maintains a running estimate of the mean of the signals, produced over a period of 5 ms, which is below the stationarity threshold for speech signals. This is in accordance with the recommendations in the paper of Matsuoka *et al.* [51], and was also shown experimentally in this study to produce better results than calculating the mean over a longer period. Consequently, during the accumulation of the average deltas, the mean of the signal varied considerably and thus the values substituted may have been mismatched to the current data and may not have resulted in a reduction in the non-stationarity of the sources.

Finally, further corroborating evidence was the increased convergence time of the mixed

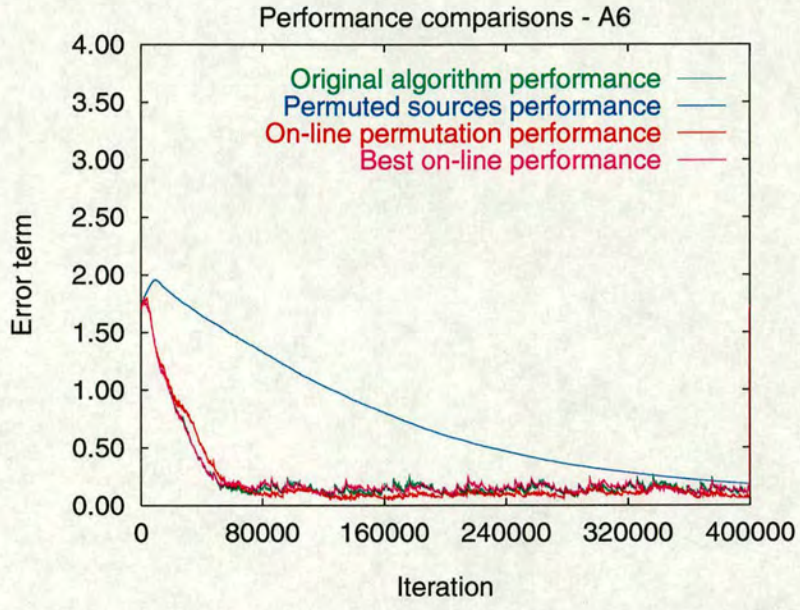


(a) Original, permuted and on-line performances for **A1** mixtures

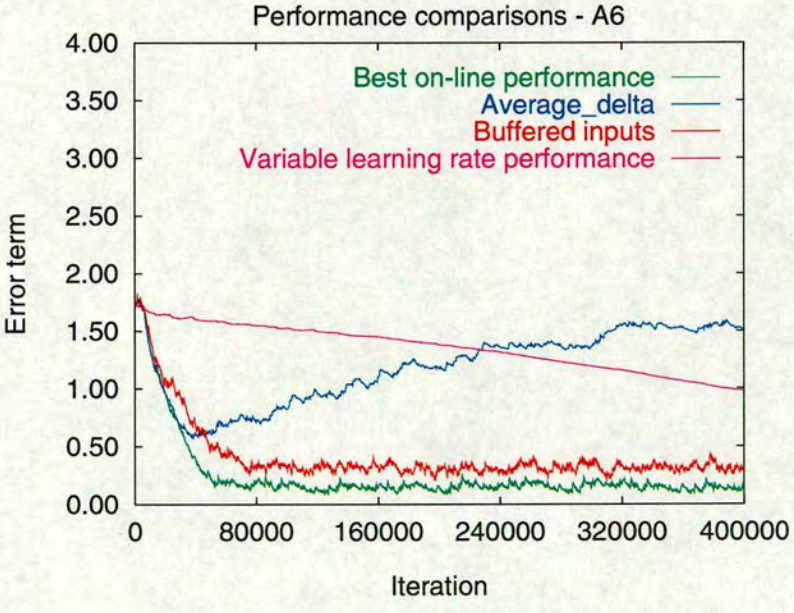


(b) Average delta, buffered inputs and variable learning rate performances for **A1** mixtures

Figure 5.22: *Separation results from the Matsuoka, Kawamoto and Ohya network for mix* **A1**



(a) Original, permuted and on-line performances for **A6** mixtures



(b) Average delta, buffered inputs and variable learning rate performances for **A6** mixtures

Figure 5.23: *Separation results from the Matsuoka, Kawamoto and Ohya network for mix* **A6**

permuted sources. As the permutation eliminated the non-stationarities of the signals, the algorithm took a particularly long time to converge — in the case of the most well conditioned system, it only just reached the level of the unprocessed sources within the duration of the experiment.

In summary, the effectiveness of the alternative update strategies was confirmed, but their application to this particular separating system was not appropriate, due to the manner in which the system operates. Consequently, the modifications to the original algorithm resulted in reduced separating performance. Overall, the original Matsuoka *et al.* system performed very well for the types of non-stationary signals (*i.e.* speech signals) considered in this study. However, it does not conform to many of the desirable properties of the other entropy-based systems, such as equivariance, and consequently cannot make use of other documented developments or techniques based around these.

5.11.4 Pre-filtering

The pre-filtering approach, unlike the network due to Matsuoka *et al.*, does not depend on the non-stationarities of the signals under investigation. It attempts to reduce the effects of non-stationary variances within the sources by applying a low-pass filter to the input signals. This reduces the degree of variance in the signal amplitude, leading to smaller disturbances from the convergence path of the network. However, it does not eliminate them completely, and consequently this approach should benefit from the adaptations being investigated.

Slight variations from the previous experimental setup were required to enable the tests to be run. Firstly, no batch-mode update was implemented due to the requirement of adapting the learning rate parameter η at each iteration, as specified in the original paper [53]. Secondly, this rate itself was modified from, $\eta_{k+1} = \eta_k - \eta_k^2$ to $\eta_{k+1} = 0.1\left(\eta_k - 0.01\eta_k^2\right)$. Failure to do this resulted in the learning rate dropping away too quickly to be of benefit, giving poor separation results. (In earlier tests without this rapidly diminishing learning rate, the algorithm failed to reliably converge.) Finally, the convergence profiling tests identified that the simulations need only be run for 160 000 iterations, which, like the natural gradient version of the infomax algorithm upon which it is based, is low compared to that of the other algorithms. This was again due to the accelerated convergence offered by the natural gradient optimisations.

After the initial spread of experiments had been run, it was found that even with the careful

parameter selection made for the learning rate, and length of simulation, the matrix inversion in the learning rule often caused the simulation not to complete. This was due to the matrix being inverted becoming ill-conditioned (near singular) and thus difficult to invert.

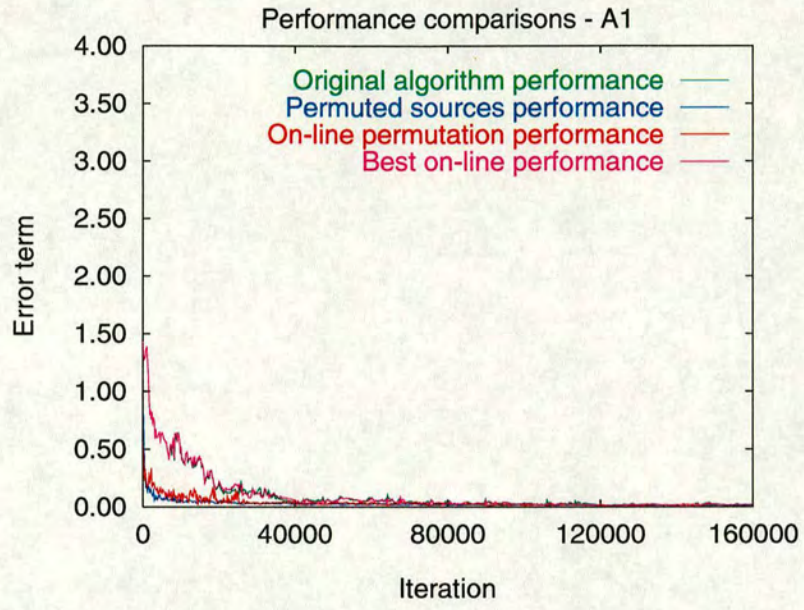
Consequently, there were insufficient results generated to carry out a fully comparable analysis as had been performed for the other algorithms, and thus the following comments are based solely on the averaged results from the subset of runs that ran to completion.

The results from the various experiments are summarised, in Figures 5.24 and 5.25, as for the previous algorithms. The graphs show that there was no major improvement in performance for any of the proposed modifications, except that of the variable learning rate and the on-line permutation. Even the variable learning rate modification appeared beneficial for only the better conditioned mixing matrices, where it offered a faster convergence rate, but no improvement in final separation performance.

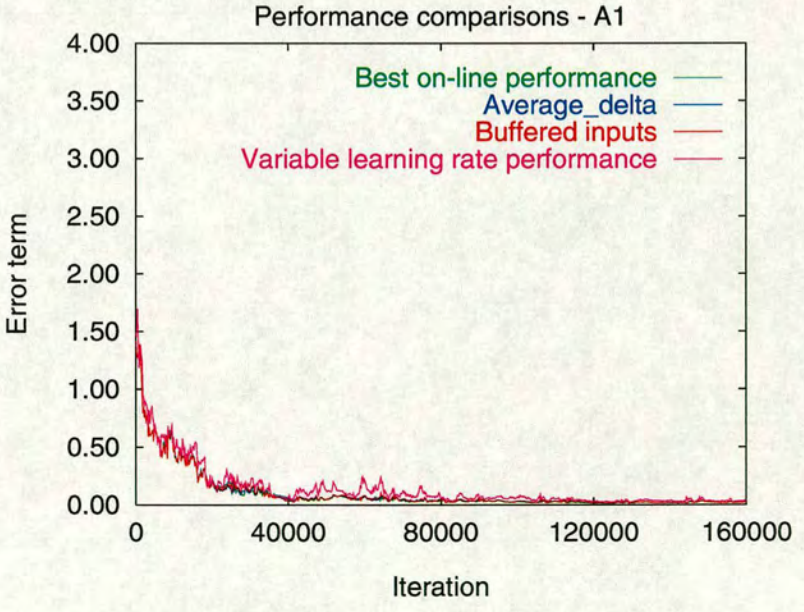
The lack of notable improvement for the basic on-line silence removal process was not particularly surprising, since the base-line performance of the pre-filtering technique is already very good and it would be difficult for any technique to improve on this. Perhaps more importantly, the pre-filtering approach itself works in a manner not dissimilar to that of the silence removal, reducing the non-stationarity of the variance of the signal. Consequently, the areas of the signal that would normally be deemed non-stationary may have already been modified in such a way that fewer of them still met the assessment criteria. Thus there would be less change to the signals and less improvement to the separation performance. It also explains why the average delta and buffered input extensions did not aid the separation in the experimentation.

In this study, the optimal silence removal parameters for the non-stationarity reduction were found to be a duration of 0.1 seconds and a threshold level of 5.00%, although the effect of average- and buffer-lengths were virtually impossible to distinguish over the range of values tried, and provided no notable improvement.

The variable learning rate method indicated some improvement because there were fewer drops to low signal amplitudes due to the improved stationarity of the variance. Thus the mean ratio of the energy of the signal buffer to that of the threshold was greater, resulting in a higher mean learning rate overall. The on-line permutation results were as expected, since the permutation had, as before, resulted in still more stationary characteristics and enabled the weight set to

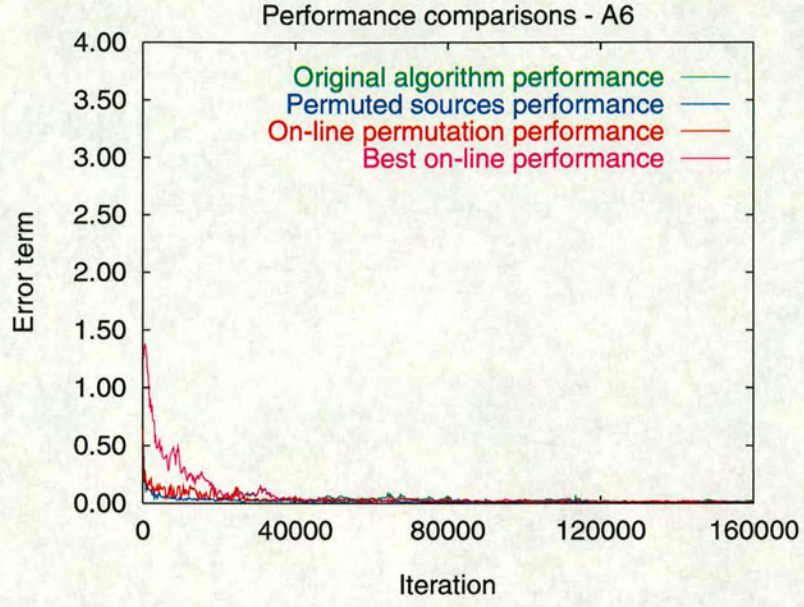


(a) Original, permuted and on-line performances for **A1** mixtures

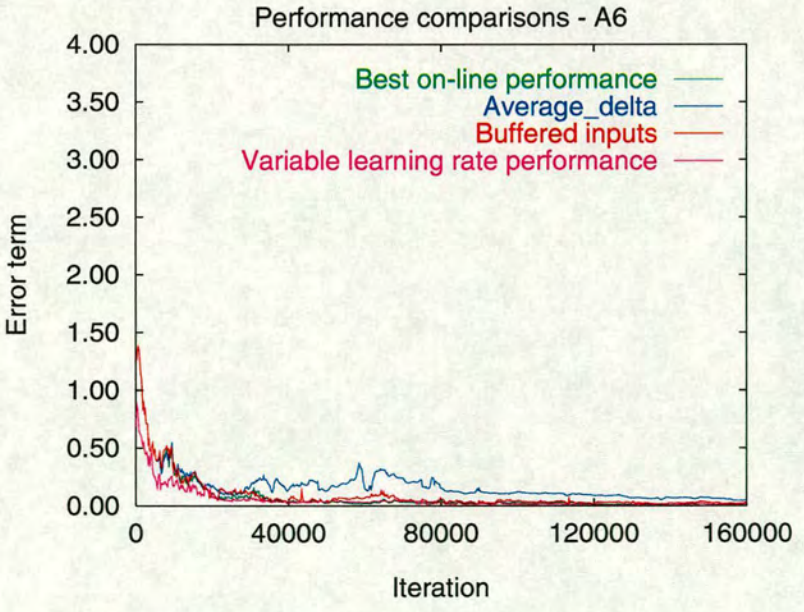


(b) Average delta, buffered inputs and variable learning rate performances for **A1** mixtures

Figure 5.24: *Separation results from the pre-filtering network for mix* **A1**



(a) Original, permuted and on-line performances for **A6** mixtures



(b) Average delta, buffered inputs and variable learning rate performances for **A6** mixtures

Figure 5.25: *Separation results from the pre-filtering network for mix* **A6**

converge more rapidly.

Overall, the change in performance offered by the proposed modifications was small. Only the variable learning rate and the on-line permutation systems provided improvement worthy of further investigation and development but only if the instability problems can be overcome.

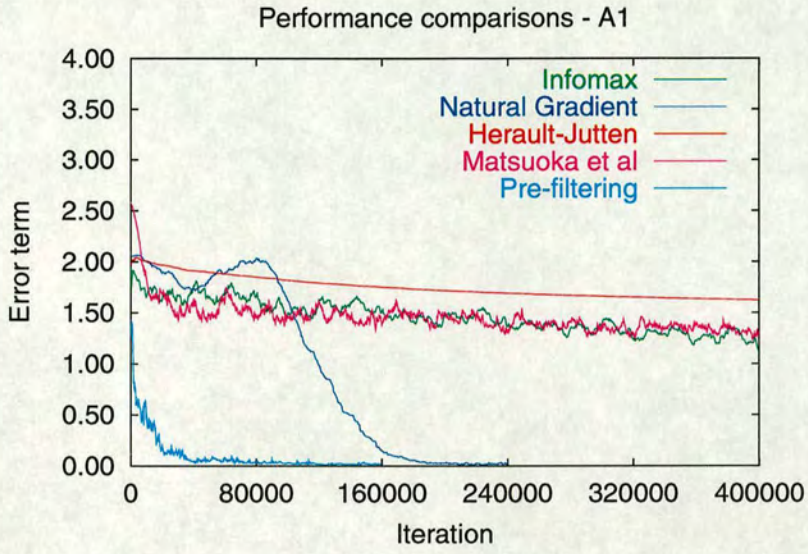
5.11.5 Relative performance considerations

By way of final comparison, the graphs in Figures 5.26 and 5.27 summarise the difference in performance offered by any of the alternative update strategies over the original, for each of the algorithms considered. (Note : corresponding traces are in the same colour for each graph.) The on-line permutation results are excluded from the comparison since they do not make use of the non-stationarity reduction techniques, and in all but the Matsuoka *et al.* case, the results are very close to those of the off-line permuted sources. It should be noted that the natural gradient results were terminated after 240 000 iterations and the pre-filtering results after 160 000 iterations, due to the lack of any change in their performance after this time. The trace for the Hérault-Jutten algorithm in the **A6** cases could be mirrored around the $y = 2$ line for a better performance comparison — doing so would yield results comparable with the others for the original algorithm, and slightly better than that of Matsuoka's system in the comparison of alternative strategies.

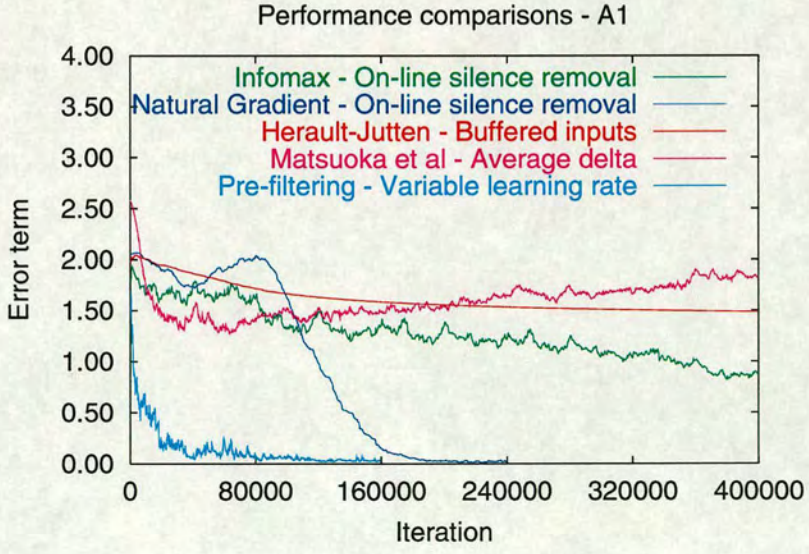
The modifications appeared to offer some marginal improvement to the performance of the Hérault-Jutten system. However, for all of the other algorithms considered, the best performance of the modified systems did not match that of the unmodified algorithm. The best performances from the unmodified algorithms was achieved by the natural gradient algorithm and that of Barros & Ohnishi's pre-filtering system, although this latter system was found to be unstable.

5.12 Areas for further investigation

From the wide ranging experimentation, a number of avenues for possible future investigation were raised, which may have potential for further advancing some of the results presented in this chapter :

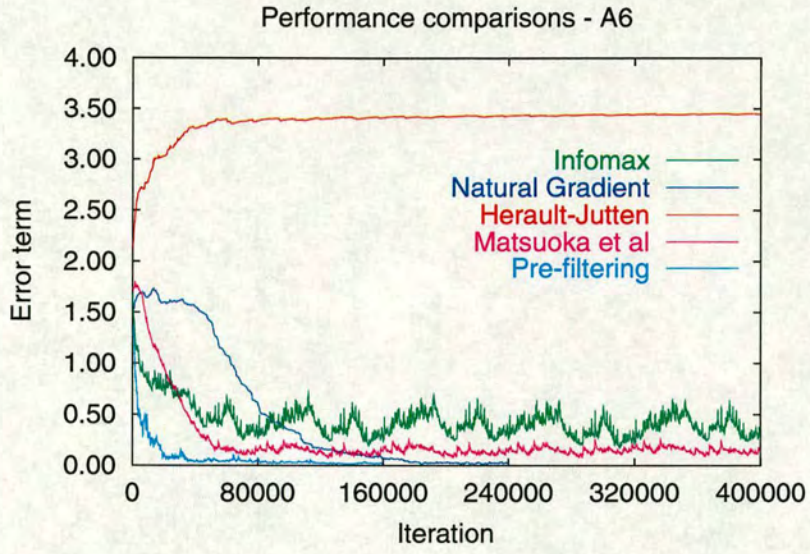


(a) Original performances for **A1** mixtures

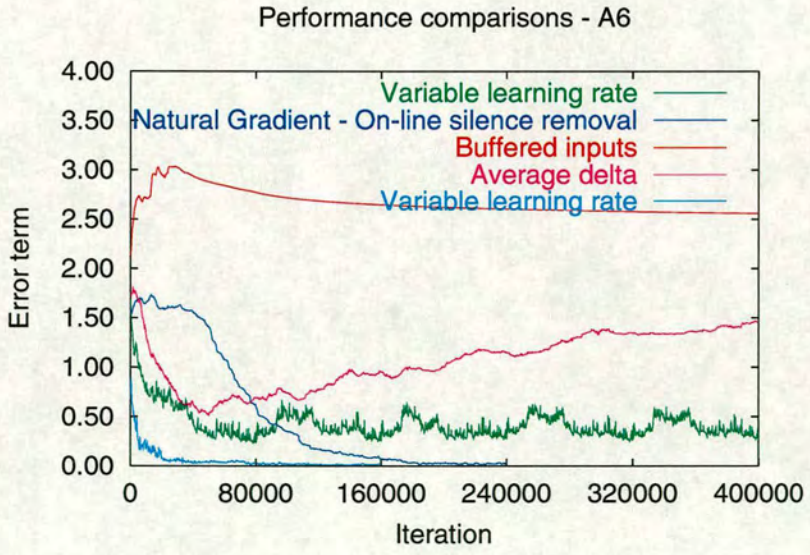


(b) Best performances for **A1** mixtures

Figure 5.26: *Comparison of the original and best performances for mix* **A1**



(a) Original performances for **A6** mixtures



(b) Best performances for **A6** mixtures

Figure 5.27: *Comparison of the original and best performances for mix* **A6**

- In the buffered inputs approach, identification of when the buffer contains data suitable for storing. In the experimentation carried out here, basing this on the energy of the buffer did not guarantee stationarity of the signals' characteristics over the duration of the buffer.
- For both the *buffered inputs* and *permuted inputs* techniques, alternative sample selection strategies may yield better performance and change the determined optimal buffer size.
- Testing the performance improvement and tracking potential offered by the use of the on-line silence removal techniques for dynamically changing mixtures of signals.

Further testing of the modifications to the other algorithms considered in Section 5.11 investigating the performance when extended to larger networks in these cases should also be undertaken.

5.13 Summary

This chapter investigated the on-line use of the silence removal techniques, presented previously, as a means of non-stationarity assessment. The original information-maximisation-based algorithm was modified to incorporate the silence removal process, and subsequently used to monitor the outputs and determine whether or not to update the network. This approach was found to offer limited performance improvements under certain conditions, at the expense of a slightly increased convergence time. Further techniques aimed at permitting updates to take place even during periods of identified non-stationarity were also investigated, including use of average updates, buffered inputs and variable learning rates. The effect of permuting the inputs was also studied. Of all the techniques considered, only the variable learning rate was found to offer any significant improvement in performance, and only for the most difficult mixtures.

Comparisons of the use of the on-line non-stationarity reduction techniques with other separating methods were also investigated — a Natural Gradient variant of the Infomax algorithm, the Hérault and Jutten algorithm, Matsuoka, Ohya and Kawamoto's algorithm and the pre-filtering techniques proposed by Barros and Ohnishi. The effect of the modifications on the performance of the Hérault-Jutten algorithm was small, offering relatively little improvement. They were also shown to have a detrimental effect on the performance of

Matsuoka *et al.*'s system due to its reliance on the non-stationarities (which demonstrated the effectiveness of the techniques at reducing the non-stationarity of the signals), and appeared to offer only slight benefit to the pre-filtering system.

Chapter 6
Non-Stationarity Reduction and Convolutive Mixtures

Following on from the work in Chapters 4 and 5, this chapter extends the investigation into the use of silence removal as a means of non-stationarity reduction, when applied to convolutive mixtures of signals. Convolutive mixing offers a more realistic model of real-world mixing as it takes into account combinations of potentially time-delayed echoes of the signals. Different techniques exist for the separation of convolutively mixed signals — some of these are carried out in the time domain, and others in the frequency domain. The effectiveness of the silence removal approach in both domains is considered.

6.1 Outline of the research

The instantaneous mixture cases considered in Chapters 4 and 5 form a logical starting point for the investigation of the applicability of the non-stationarity reduction techniques. However, to assess how well separation techniques would perform in a real-world application many additional factors should also be considered, such as convolutive mixing effects, the presence or absence of noise, and the stationarity of the mixing process, to name but three.

This study addresses the first of these — convolutive mixing effects. As previously described in Section 3.1 convolutive mixing means that the signals are no longer merely mixed with one another on a sample by sample basis, but also with time-delayed echoes introduced by filtering effects of the propagation environment, such as room characteristics. The aim of the separation process is now modified as it must also filter the signals to remove these echoes, thus deconvolving them. This is a considerably more complicated task than separation alone, particularly as the two processes must be carried out simultaneously — if the problem could be split into its two parts and one carried out before the other, the existing techniques for dealing with the individual problems could be applied. However, this is not the case, and hence solutions that address both issues jointly are required. These techniques may carry out the

processing in either the time domain, as in the instantaneous case, or move to the frequency domain in an attempt to reduce the computational load of the processing required to solve problems involving either long filters, or large numbers of signals.

The work presented in Chapters 4 and 5 is further developed to allow it to be applied to these convolutive mixing cases. This research assesses the impact of silence removal as a means of non-stationarity reduction on the performance of a separation and deconvolution solution, described below, when applied in both the time domain and frequency domain. Some aspects of this work are similar to the non-permanent learning reported by Nguyen & Jutten [9] and to the intermittent adaptation investigated by Van Gerven & Van Compernolle [6–8] described in Section 6.6.2.

6.2 Separation algorithms considered

Following the line of investigation developed in Chapters 4 and 5, an information-maximisation-based algorithm was sought that was capable of dealing with convolutive mixtures. Bell & Sejnowski's original paper [10] describes the use of their infomax algorithm to achieve this. This approach was further extended by other researchers, notably Torkkola [2–5], Lee, Bell & Lambert [90], Lambert & Nikias [102], and Westner [122]. Preliminary experiments with several of these algorithms confirmed results reported by their respective authors that the systems do not perform as well on signals that exhibit temporal correlations, such as speech signals. This poor performance is due to the algorithms attempting to decorrelate successive samples from the same output, as well as those from different outputs, which is the desired behaviour [103]. Some of these infomax-based algorithms have been applied successfully to separation and deconvolution problems in the field of digital communications [90], where the signals of interest are not temporally correlated.

Other approaches have been taken to solve this problem, including those based around the use of second-order statistics [108, 109, 123]. While it has been argued by some researchers [10], that these may be insufficient, the authors of such systems have demonstrated their successful use under a variety of conditions. Parra & Spence [105, 106] explain that if the signals' characteristics are non-stationary, then second-order statistics decorrelated at multiple time delays are indeed sufficient for the separation. One such algorithm, that makes use of these decorrelations at multiple time delays in the frequency domain, is the CoBliSS algorithm

by Schobben & Sommens [108, 123] (see Section 3.5.2.5). Initial experimentation with an implementation of this algorithm showed that it ran reliably in the framework to be used, and gave better results than the information-maximisation algorithms tested. Consequently, CoBliSS was selected for use in the convolutive investigations.

6.3 Modifications to the CoBliSS algorithm

The CoBliSS code required several modifications to be made to enable it to fit into the framework used to run the spread of simulations. It also required modification to incorporate the on-line form of the silence assessment and the non-stationarity reduction code. Both the silence assessment and the non-stationarity code *per se* also required some modification to accommodate the way the CoBliSS code dealt with the output data and the weight updates.

Two algorithmic modifications were made to the CoBliSS code, one applying the silence assessment to the outputs in the time domain and the other applying it to the frequency domain version of the outputs. This allowed a comparison of the effectiveness of the technique in the two domains.

6.3.1 Time domain

Although the time domain assessment of the output signals is the same process as was used for the instantaneous cases in the experiments of Chapters 4 and 5, the application of this technique required minor modifications as the CoBliSS code processes its data on a block-by-block basis. Hence, in order to achieve as similar an effect as possible to the instantaneous case, each block of output data was then assessed on a sample by sample basis, just as they had been in Chapter 5. This produced a block of assessments for each block of output. Each block of assessments was then analysed, and only if none of the assessments within the block reported any silence was the update to the frequency domain power cross-correlation matrix carried out. The frequency domain weight updates are derived from this cross-correlation matrix and the inputs, and then the time domain weights can be calculated from the frequency domain weights by means of an inverse Fast Fourier Transform (IFFT).

The reason that this approach required none of the assessments within a block to identify a period of silence before performing the update stemmed from two considerations. The first

consideration was that the original (instantaneous) assessments had required all of the outputs to be non-silent before updating the network — therefore demanding the same of each sample of the block assessment was consistent with this approach. The second consideration was the manner in which the updates are calculated by the algorithm — the update calculations are based on the frequency domain cross power spectra of the estimated outputs. There is no mapping that yields an update for a single weight in the time domain from a frequency domain weight, or *vice versa*.

One possible method of dealing with this difficulty of identifying which weights are to be updated would have been to count the number of time domain assessments that reported no silence, and use this count as a proportion of the whole block to scale all of the updates. However, since this approach still does not provide control over individual weights, and may adversely impact weights that would otherwise be unaffected, it is not considered a particularly satisfactory solution, and therefore was not pursued.

Requiring all of the assessments to report no silence before carrying out an update also differs from the approach taken by Van Gerven and Van Compernolle in their work on intermittent adaptation [6–8]. There, they permitted the update of either channel independently of the other. However, since the system they used to investigate this approach had only two channels, the conditions under which either channel could be updated were relatively simple to determine. In the experimentation carried out in this thesis, the approach taken was designed to be extendable to any number of signals without significant increase in complexity, although it may not be able to take advantage of updates to individual channels.

6.3.2 Frequency domain

There is good reason to consider the application of the non-stationarity reduction in the frequency domain separately from that in the time domain. This is because each frequency of interest can be considered independently of the others, and the assessment of whether or not to update its associated weight carried out accordingly. This would allow individual weights corresponding to particular frequency bins to adapt, whilst suppressing the update of others whose channels were deemed to be silent. This should assist the overall separation of the system, and extends the original work of Van Gerven and Van Compernolle [6–8] which considered only the time domain assessment of the output signals.

The principle behind applying the non-stationarity reduction techniques in the frequency domain is the same as that for the time domain — to attempt to limit the degree of variation of the updates. This variation is caused by rapidly changing amplitude of the signals being assessed, and can be reduced by the application of silence removal. Now that the signals under assessment are the successive values of the frequency components of the output signals, some further modifications to the assessment techniques and their parameters are required.

In the time domain, the assessment consisted of evaluating the energy of the signal over a buffer of specified duration, and comparing this to a threshold level. The energy value for the buffer was updated as successive samples of the output signals were added into the buffer, and the oldest values removed. To transform the signal into the frequency domain, N time domain samples were used to carry out a Fast Fourier Transform (FFT) into M frequency bins. Thus the energy of the N samples was spread over the M bins. However, the FFT operation used does not normalise the values in its forward operation, only in its inverse, and hence the forward values were M times larger than they should be. These two factors of M — one from the forward FFT operation, and the other from the distribution of the signal energy across the M taps — cancel each other out. This means that the same threshold levels that were used in the time domain assessment can also be used, unaltered, in the frequency domain assessment.

The duration parameter, which corresponds to the length of the assessment buffer, must also be considered. The aspect of interest is the relative value of this parameter compared to the fixed number of samples, N, to which the FFT operation is applied. When the duration corresponds to a number of samples less than N, only the current FFT output values need be considered. In this case, the energy value for each frequency bin can be scaled by the ratio $\frac{duration}{N}$, to compensate for the greater amount of energy present as a result of the larger number of data points transformed. When the duration corresponds to a number of samples greater than N, successive FFT values must be buffered for each frequency bin, and again the energy value scaled to compensate for the mismatch in the number of samples present. The number of samples that must be buffered can be determined from Equation 6.1.

$$Q = \text{floor}\left(\frac{duration}{N}\right) \tag{6.1}$$

The scaling factor required is then given by $\frac{duration}{QN}$. This same scaling can also be applied to the former case, $duration < N$ if an additional provision is made that Q takes a minimum value of 1.

In both cases, rather than scaling the energy values, the threshold level could have been scaled by the inverse of the prescribed ratio. One reason for not taking this approach was the possibility of a future extension to the work that would apply a shaping filter to the values of the frequency bins. This extension would either require the creation of M separate thresholds which would be individually scaled, or could factor it into the current scaling operation. Such shaping filters could be used in non-blind situations to provide additional information about the expected frequency spectra of the signals being separated.

Having considered the necessary changes to both the threshold and duration parameters, the resulting assessment is as close as is possible to its time domain equivalent. As before, the choice of which elements of the frequency domain power cross-correlation matrix (and thus which weights) to update, according to which frequency bin outputs pass the silence assessment, needs to be determined. Again, it would be possible to allow each weight to adapt individually, but for consistency it was decided to study two approaches :

1. Require all elements from all frequency bins in all output channels to have satisfied the silence assessment
2. Require corresponding frequency bins from each output channel to satisfy the silence assessment

The former approach would allow a direct comparison with the time domain version described above in Section 6.3.1. The latter approach exploits the greater flexibility by permitting individual frequency bins to update according to the assessment of that frequency component of the signals. The two approaches will be referred to as **FD_all** and **FD_select** respectively.

6.4 Experimental setup

The experimental setup used was kept as similar as was possible to that used for the instantaneous mixture simulations. This meant that the simulations could be run and the results collected and assessed in a similar manner. Certain changes were required, however, to accommodate the differences between the instantaneous and the convolutive experiments. These are described below.

6.4.1 Convolutive mixing

To assess the applicability of the non-stationarity reduction techniques to real-world situations, it was decided to evaluate their performance on convolutively mixed signals. As described in Section 3.1, this means that the two-dimensional mixing matrix of the instantaneous case is replaced by a matrix of filters — effectively a three-dimensional array of values. The resulting effect of this is that the signals are not just mixed, but are also convolved with their time-delayed echoes.

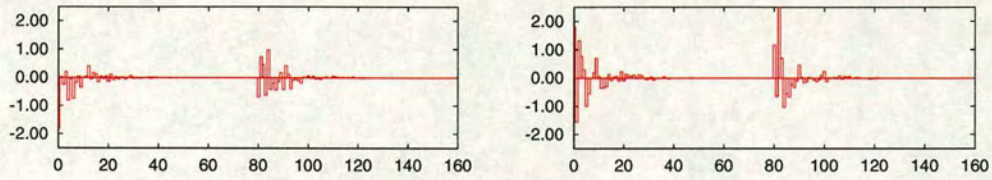
In determining the filters to be used for these experiments, consideration was given to the factors of the techniques to be assessed. Since the silence removal technique used to achieve the non-stationarity reduction is dependent on a duration parameter, filters were selected that were appropriate to the periods of interest. It was also appropriate that the filters have real-world characteristics, such as a typical room-impulse response. Leading delay taps were not considered here.

It was decided to investigate the effect of the silence removal on signals that had been filtered by a system with a room-like impulse that incorporated a large echo component. The period between the echos was varied, and for different filters was selected to be either above or below the stationarity threshold for speech. Four sets of such filters were used, each of different length, as shown below in Figures 6.1, 6.2, 6.3 and 6.4, and the corresponding periods between the echoes are given in Table 6.1.

Filter set	Echo length (ms)	Taps separation	Total taps
F1	10	80	160
F2	20	160	320
F3	30	280	560
F4	100	800	1600

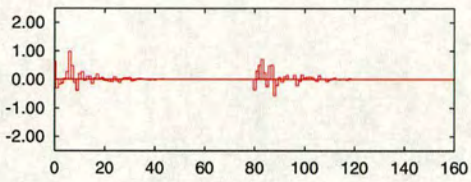
Table 6.1: *Convolutive filter sets and echo delays*

Each set comprised four filters, to allow a two by two mixing and filtering operation to be performed. The individual filters were generated using Matlab [124], based on a negative exponential distribution that dropped by 40 dB over the duration of the filter. These filters

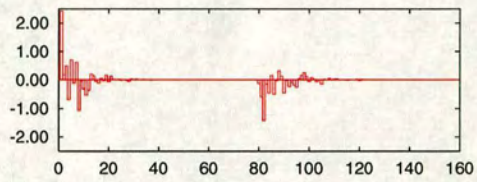


(a) F1 : $filter_{00}$

(b) F1 : $filter_{01}$

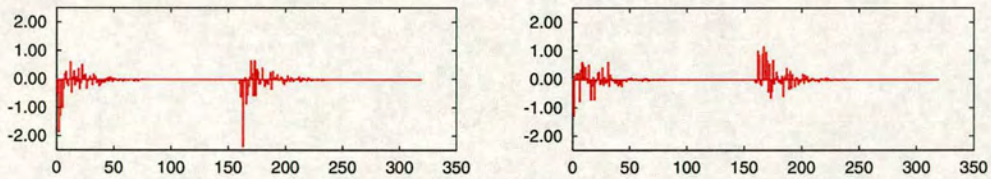


(c) F1 : $filter_{10}$



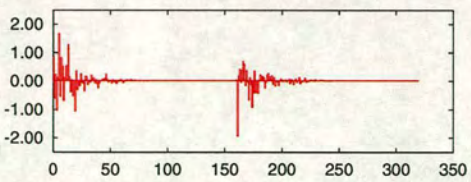
(d) F1 : $filter_{11}$

Figure 6.1: *Convolutive mixing filters - F1*

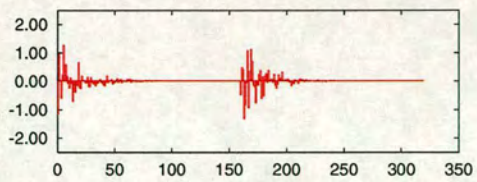


(a) F2 : $filter_{00}$

(b) F2 : $filter_{01}$

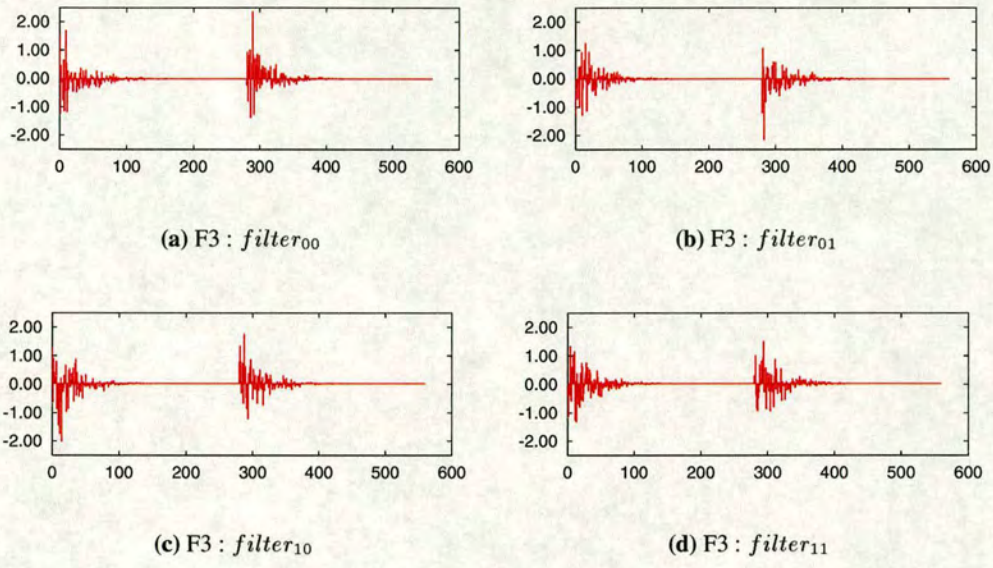


(c) F2 : $filter_{10}$



(d) F2 : $filter_{11}$

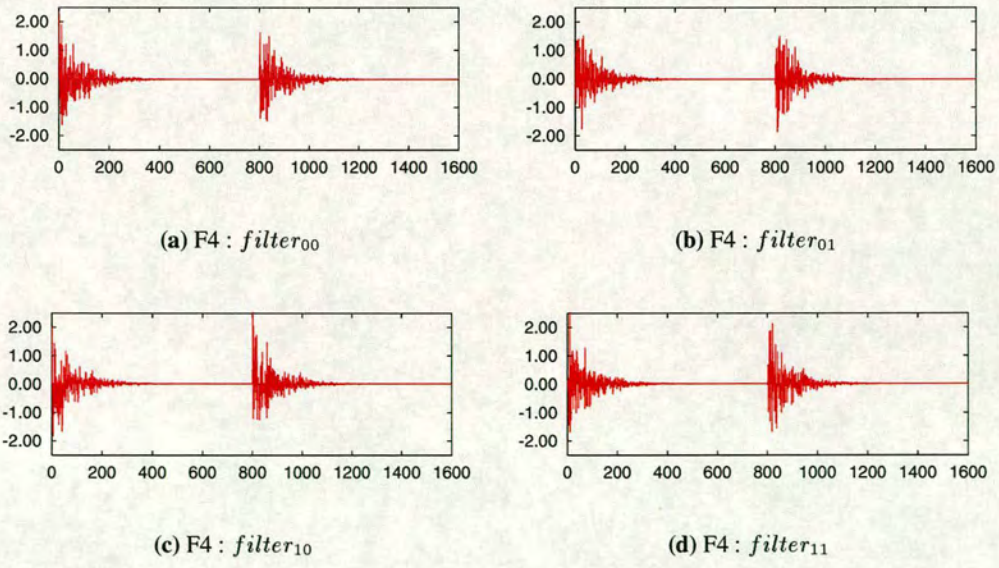
Figure 6.2: *Convolutive mixing filters - F2*



(a) F3 : $filter_{00}$ (b) F3 : $filter_{01}$

(c) F3 : $filter_{10}$ (d) F3 : $filter_{11}$

Figure 6.3: *Convolutive mixing filters - F3*



(a) F4 : $filter_{00}$ (b) F4 : $filter_{01}$

(c) F4 : $filter_{10}$ (d) F4 : $filter_{11}$

Figure 6.4: *Convolutive mixing filters - F4*

were compared to randomly-generated filters based on exponentially decaying Cauchy noise, produced by publicly available code provided by Smaragdis [125]. The filter characteristics of both sets were found to be very similar, with the exception of the additional echo in the custom designed sets. The custom defined filters were selected for use due to the additional control available in their generation.

These sets of filters are relatively complex, and it is well known that longer, more dense filters are more difficult to learn than simpler ones. However, this complexity is a necessary result of the desired characteristics of the filters for this investigation.

6.4.2 Performance metric

The performance metric used for these experiments was based on the Amari *et al.* metric [58] used in the previous chapters. Other metrics for the evaluation of separation and deconvolution systems do exist, but many of these assess the output signals [105, 106, 126], and not the separating system itself. In their investigation of digital communications systems, Lambert & Nikias [102] make use of measures of the multichannel row and column Inter-Symbol Interference (ISI), but this could equally well be defined in terms of multichannel Signal-to-Interference Ratio (SIR), thereby removing the implied use with digital signals. Individual SIR measures can give an assessment of the performance of a single filter in the separating and deconvolving system, but to gain an performance assessment for the system as a whole, such measures must be combined in some way. By comparison, the Amari *et al.* metric provides this global measure directly.

Although the Amari *et al.* metric was originally defined for two-dimensional mixing matrices for the instantaneous case, the principle on which it works was extended to accommodate three-dimensional arrays, allowing it to be used to assess the performance of separation and deconvolution systems. In the instantaneous mixing case, the product of the mixing matrix and the separating matrix should yield a scaled, permuted identity matrix. For the convolutive case, the convolution of the set of mixing filters and the set of separating filters should yield a scaled, permuted set of impulse response filters. The SIRs for each of these resulting filters can be calculated, and the two-dimensional matrix of SIR values will have the property that for each row or column there should be only one non-zero value. The formula for the Amari *et al.* metric (see Section 4.5.1) can then be applied to these values to give an overall assessment of the system. This is the metric used in all of the simulations described below.

6.4.3 Weight set initialisation

In the original CoBliSS algorithm, the weight sets are initialised to the inverse of the cross-correlation matrix of the power spectra. This will be a decorrelating weight set, which should thus start the evolution of the weights from a state close to the separating and deconvolving solution. However, it is not guaranteed that this decorrelating initialisation will necessarily be close to the desired separating and deconvolving state.

In this study, such a fixed initialisation would have conflicted with the weight initialisation used to provide an independent start point in the multiple replicates required for statistical analysis. In order to preserve this option, preliminary experiments were run to compare the performance of the separation and deconvolution experiments when started with weight sets initialised to values other than that prescribed by the algorithm. The results showed only a small difference in performance after a fixed number of iterations, suggesting that the benefit of the initialising decorrelation weight set was not great. An analysis of the end points of the simulations showed that the differences identified were not statistically significant at the 5% level.

Consequently, the fixed initialisation procedure was replaced by the scheme used previously — that of using a set of randomly generated weight sets, now extended to accommodate the additional dimension required. This enabled the simulations to be run using a similar framework to that used previously, the independent start points providing additional replicates for the statistical analysis of the results.

6.4.4 Statistical analysis

The experimentation was designed to facilitate statistical analysis of the results generated. Initial analysis of the data from the simulations carried out for the work in Chapters 4 and 5 showed that the significance levels of the analyses were very high due to the high precision of the results and the very large number of simulations carried out. For these convolutive studies, the number of replicates run was reduced on the advice of the statistician consulted [120]. Lowering the number of replicates from thirty to sixteen (four different initial weight sets at each of four different time offsets into the signals) increased the Standard Error of Difference (SED) and the Least Significant Difference (LSD), but within acceptable limits and without losing any of the desired detail.

6.5 Experimentation and results

The same source signals as had previously been used to create the instantaneous mixtures were used with the filters described above in Section 6.4.1 to generate the convolutive mixtures for the input data.

Convergence profiling of the original algorithm was carried out as before (see Section 5.2.3, with the distribution of the end points of the simulations being assessed using the Coefficient of Variation (CV). After running the simulations for up to fifty passes through the input data (4800000 data points), the CV was still found to vary for each of the filters, suggesting that the systems had not yet converged to their final values. It was decided to curtail the experiments after 2400000 data points, due to the computing time it would have taken to carry out the experimentation beyond this level. This value was chosen from the range of simulation lengths considered as it had the overall lowest set of CVs for all four filter sets.

Due to the mechanism for calculating updates, each simulation that uses a differing length of the filter sets processes a different number of blocks of the data for the same total amount of data. Since performance assessments are calculated once per block, this results in different numbers of assessments being carried out for the different filters. The graphs presented below have been rescaled to account for this fact, and are plotted against the individual sample index of the data. The rescaling accounts for the apparent different density of the traces for the individual filter sets on the graphs. The different amounts of processing also meant that the simulations involving the longer filters took considerably longer to run.

Results for the different experiments are presented below, and discussed in Section 6.6.

6.5.1 Original algorithm results

The results of the original algorithm averaged over all of the different preload sets and time offsets for each of the four sets of mixing filters, $\mathbf{F1}$, $\mathbf{F2}$, $\mathbf{F3}$ and $\mathbf{F4}$, are shown in Figure 6.5. The experiments showed that none of the simulations performed particularly well, although a trend can be seen in the performance over the different filters — the trend suggests a slight reduction of separation performance with increasing filter length and is as expected, since the longer filters are more complex, and hence more difficult to invert. According to the statistical analysis (see Table 6.2), these differences in performance are not significant at the 5% level.

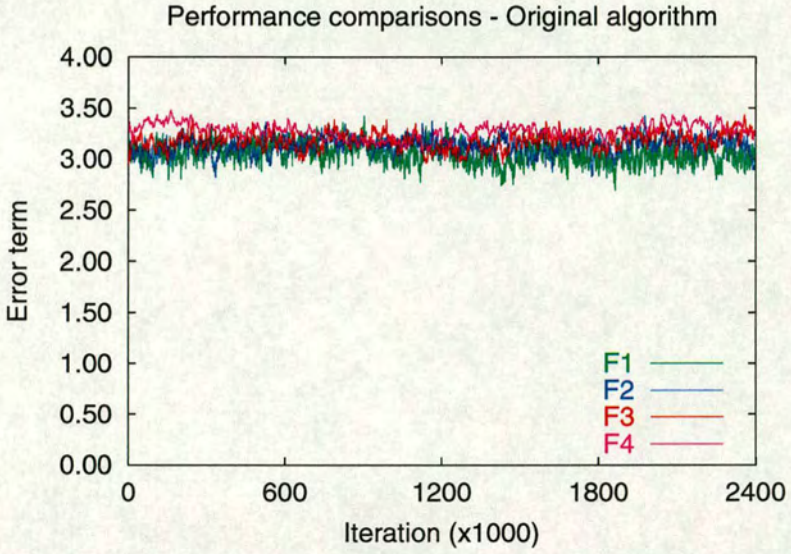
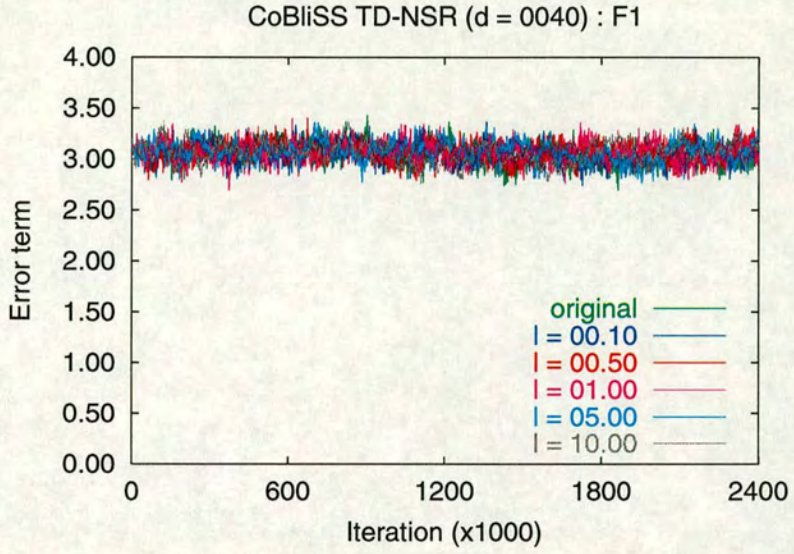


Figure 6.5: *Original CoBliSS algorithm performance*

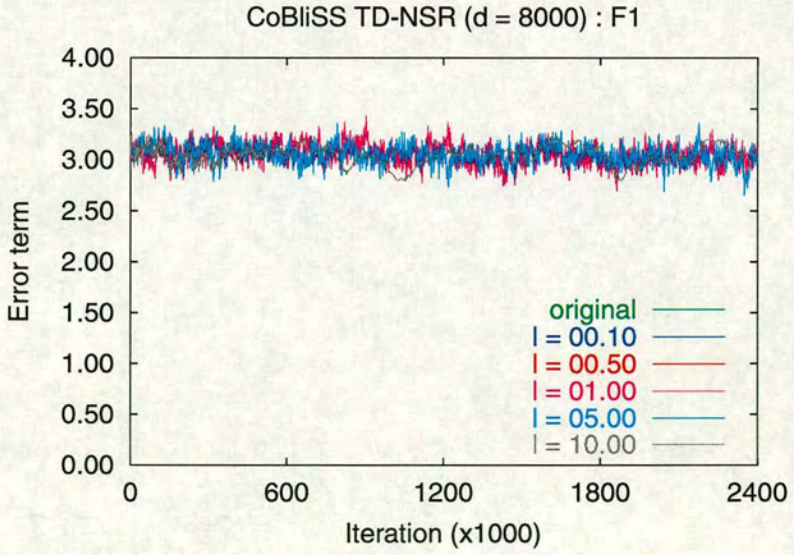
Filter set	F1	F2	F3	F4	Reps
Performance	3.078	3.181	3.275	3.285	16

D.F. = 60 L.S.D. (5%) = 0.1912

Table 6.2: *Separation performance by filter set*

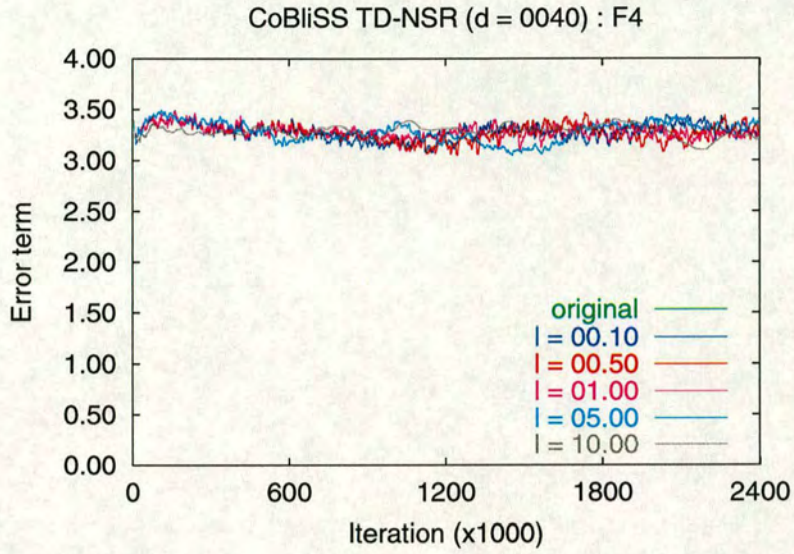


(a) F1 : d = 0.005 s

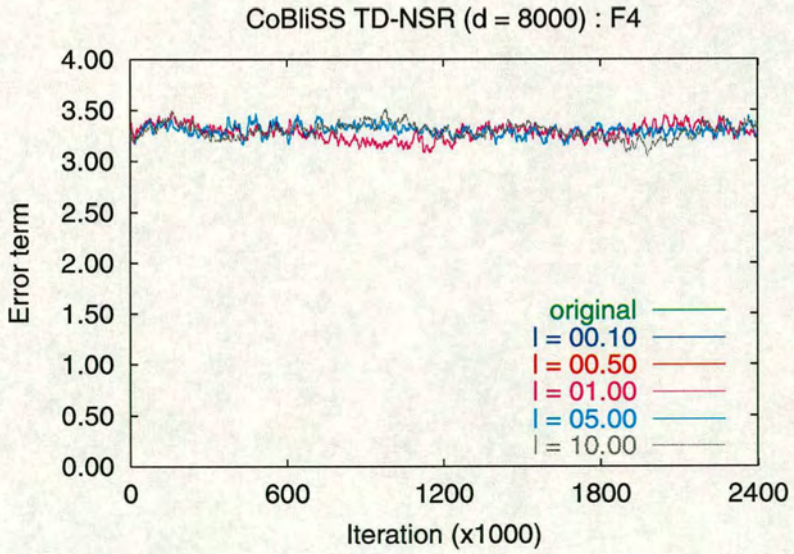


(b) F1 : d = 1.000 s

Figure 6.6: *Effect of time domain assessment on separation performance for* **F1**



(a) F4 : d = 0.005 s



(b) F4 : d = 1.000 s

Figure 6.7: *Effect of time domain assessment on separation performance for* **F4**

6.5.2 Time domain results

The results of the simulations using the time domain assessment of the output signals appear largely similar between the different mixtures. Consequently only the graphs for the shortest and longest filters (**F1** and **F4**) are presented. Figures 6.6 and 6.7 show the results of these experiments for the same range of parameters used in the earlier instantaneous mixture investigation — a range of different threshold levels at two fixed durations : $d = 0.005\ s$ and $d = 1.000\ s$. Many of the results for the different threshold levels at each of the fixed durations overlay one another, making them difficult to distinguish from each other and from the performance of the original (unmodified) algorithm. Examination of these results, for all possible combinations of parameters on each filter set, suggested that the best performance was achieved with a duration of $0.100\ s$ and a threshold level of 1.00%. However, analysis of this result showed that this was not significant and that the only factor in the experiments that caused any significant difference in performance was that of the filter set. (See Appendix B.)

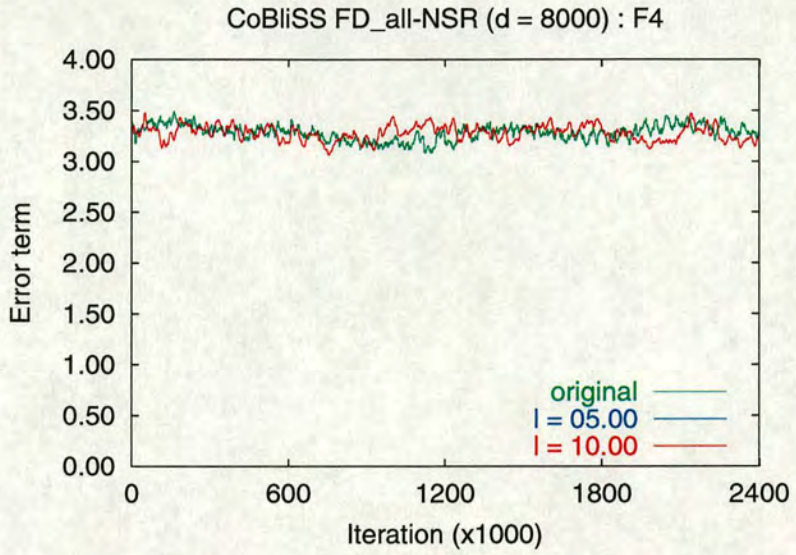
6.5.3 Frequency domain results

When running the experiments for both versions of the frequency domain assessment of the outputs, it was found that many of the simulations did not run to completion. Those that finished prematurely had aborted, often due to numerical problems in inverting matrices using presupplied library calls. Likely reasons for this are discussed in Section 6.6.3.

When using the **FD_select** approach, none of the parameter combinations tested completed sufficient runs to produce a graph using the standard averaging technique that had previously been used. From the **FD_all** experiments, the runs with parameters set at the higher values ($d = 0.500\ s, 1.000\ s$ and $l = 05.00\%, 10.00\%$) for filters **F3** and **F4** produced traces that were averaged and are shown in Figure 6.8. As with the time domain graphs, the traces for the higher threshold levels are plotted later, overlaying those at the lower levels. Due to the low number of completed runs, statistical analysis was not carried out on these data sets.

6.6 Discussion

It is possible that the poor performance of the information-maximisation algorithms initially investigated was due to the combination of at least three factors known to adversely affect



(a) F4 : d = 1.000 s

Figure 6.8: *Effect of frequency domain assessment* FD_all *on separation performance for* **F4**

the performance of such systems — the temporal correlations within the mixed speech input data, the rapidly changing amplitude of the signals and the complexity of the filters under investigation. Such infomax-based algorithms are known to perform far better on i.i.d. signals, although some success in separating speech signals has been reported using these algorithms. To enable an investigation of the application of time domain and frequency domain-based non-stationarity reduction to infomax-based systems, appropriate base-line conditions, configurations and signals would first need to be identified.

6.6.1 Original CoBliSS algorithm results

The performance of the original CoBliSS algorithm on the systems investigated was poor in comparison to those achieved during the instantaneous separation experiments of Chapters 4 and 5. This was disappointing as good results have been reported from real-world experiments, and for filter lengths of up to 512 taps [108]. Use of this algorithm was discussed in [126] as the starting point for a wider investigation of the behaviour of separating algorithms generally, under a range of mixing and filtering conditions. Despite it not being so susceptible to the temporal correlations within the input signals, its performance in the preliminary experiments

was only marginally better than that of the information-maximisation-based algorithms of Lee, Bell & Lambert [90], and of Lambert & Nikias [102].

This poor performance, even for the shortest filter set **F1** of 320 taps, was attributed to the complexity of the filters used in the creation of the input signals. Whilst it would have been possible to use more simple filters, those used were designed specifically to illustrate particular characteristics such as a typical real-room response and a strong echo component at differing delay intervals. It was considered desirable to incorporate the former characteristic in such investigations to allow the results to be generalisable beyond the laboratory experiments carried out. The latter characteristic was of interest with regard to the duration parameter of the non-stationarity reduction used, the intention being to investigate whether the separation performance was different when the echoes occurred at periods less than the duration of the sections of silence being removed and *vice versa* for durations on both sides of the stationarity threshold for speech.

6.6.2 Time domain assessment

The results from the experiments using the time domain assessment of the output signals to control the adaptation of the system showed little variation in their results. Although it was possible to identify values for each of the parameters that gave the best performance over the range of conditions considered, the improvements achieved with this combination were shown not to be statistically significant.

This is interesting, as Van Gerven & Van Compernolle [6–8] had previously shown their intermittent adaptation technique to be beneficial to the separation results. Intermittent adaptation uses a scaled comparison of a windowed energy estimate of each of the output signals to determine whether or not to update the filter coefficients, in a manner similar to that employed in the simulations carried out above. Nguyen Thi & Jutten [9] also comment on the beneficial effects of what they refer to as "non-permanent learning", another similar technique that only updates a filter if the energy of its corresponding outputs is above a set level, although they provided no other details of the technique.

The separating systems used in both of these other studies are also decorrelation-based, like the CoBliSS algorithm, although both Van Gerven & Van Compernolle's Symmetric Adaptive Decorrelator and Nguyen Thi & Jutten's system use higher order statistics to decorrelate the

signals from one another.

In Van Gerven & Van Compernolle's experiments, and in those reported by Nguyen Thi & Jutten, the filters considered were much shorter and simpler than those used in this study. Van Gerven [8] reported good results with a window length of 100 ms and an energy threshold ratio of 0.6, while Nguyen Thi & Jutten estimated the signal energy over a 200 ms window. Both of these window durations are within the range of the parameters investigated here, but were not found to produce significantly different results in this study. However, it should be noted that both sets of authors do comment that optimal values are likely to be signal specific.

The simulations performed for the research in this thesis use the threshold as the scaling factor for the maximum energy seen so far on that channel, rather than as a scaling factor for the current energy level of the other channels. This difference in the assessment method between the two systems would lead to differences in the decision of when to carry out an update.

It is also possible that the different results are due to the method used to determine whether or not to carry out the updates based on the output assessment. The intermittent adaptation approach updates the weights in a filter only if the ratio of the energies in the two channels being compared is greater than the specified ratio. In the approach used here, all of the assessments within a block, for all of the channels considered, must pass the assessment if an update is to be carried out. This is a distinctly different strategy, which is likely to result in far fewer updates being carried out than in the intermittent adaptation approach, which will in turn produce very different results in the final assessment of such systems.

The key reason for this different update strategy is the fact that the assessment is being carried out in the time domain, producing a decision on a sample by sample basis, while the update itself is applied to the frequency domain power cross-correlation matrix. Since there is no direct mapping between a single time indexed element and individual frequency indexed elements, it is impossible to restrict the time domain assessment results to specific frequency domain elements within the block. This will be true of any system that attempts such specific cross-domain control, and consequently an "all-or-nothing" type approach must be adopted.

It was recognised that it would be possible to use the assessments to determine whether or not to update all of the elements of a particular filter and that this would allow greater flexibility in the control for these approaches. However, to ensure consistency of approach, this possibility

was deliberately not pursued in this part of the study, as was the case in the earlier work on the instantaneous mixtures.

6.6.3 Frequency domain assessment

The poor performance of the frequency domain assessment modifications, in terms of the low number of completed simulations, meant that rigorous comparisons could not be made with either the time domain version, or with the original algorithm. The numerical problems encountered that lead to the abrupt termination of many of the simulations are believed to have occurred during the calculation of the weight updates from the frequency domain power cross-correlation matrix, which involves several matrix inversions. This suggests that the matrices being processed, each corresponding to a particular frequency bin, had become uninvertible.

A possible explanation for this, that could not be fully investigated due to time constraints, relates to the decision taken with regard to the 'forgetting factor' of the cross-correlation matrix update. This factor behaves in a similar manner to tap leakage — a standard technique in adaptive filtering, used to achieve a minimum-norm solution by reducing the magnitude of all current taps by a small amount at each iteration. The 'forgetting' occurs at the same point that the new updates to this cross-correlation matrix are added. However, in the modified version of the algorithm using the **FD_select** assessment, these updates may not be added according to the output assessment. Consequently, the values in this matrix may gradually decrease towards zero, which could have lead to the observed problems. Whilst it would have been possible to suspend the 'forgetting' in conjunction with the suspension of the update for a particular element in this approach, this was deliberately not done so that the results produced could be compared with the **FD_all** approach.

In the **FD_all** modification, both 'forgetting' and update are conditionally suspended for the whole block unless all of the assessments report sufficient energy in their frequency bin for all channels. This strategy would have allowed a direct comparison with the time domain approach which made use of the same strategy, but looking only at blocks of data for each of the output channels, rather than for each frequency bin.

It is likely that the reason that the simulations encountered this problem was that not all of the frequency bins contained sufficient energy to pass their assessment regularly enough to

maintain an accurate estimate of the true cross-correlation matrix. This would also have been the point where the shaping of the thresholds would have played a significant role, making it more likely that all of the different frequency bins pass their assessment once their expected relative levels had been taken into account.

Another possible reason for the problem was the length of the filters considered, and the block size of the data processed. The lengths of the filters involved ran from 160 taps upwards. This length corresponds, at the 8 kHz sampling frequency used, to a period of 20 ms. Since the data was regarded in blocks of twice the filter length, this means that 40 ms worth of data were being transformed at a time, even for the shortest filter set — well above the stationarity threshold for speech. Consequently, the different characteristics of the active and silent parts of the signal will be blurred, combined into one block of frequency domain values, and the periods of silence sought to control the adaptation masked within these blocks. As a result, it is possible that the desired adaptation could not be achieved in this framework.

To overcome this problem, the block length of the data being processed would need to be reduced. This approach would also require separating the handling of the input data from that of the weight data to allow the necessary short window assessment to be carried out, but then to link them back together again — possibly after the recombination of multiple blocks of input data — to permit the processing of the weights and their updates to proceed. Whilst this type of change should be possible, it may considerably complicate the algorithms.

An advantage of this approach is that if the input data has distinct 'phases' in its non-stationarity, the different characteristics of each phase may yield different information about the system, and this may be exploited once these characteristics are made visible by the reduced block length. However, an associated disadvantage is a loss of generality — this scheme requires foreknowledge of certain properties of the input data, such as the anticipated period of its non-stationarity, which is not possible in a truly blind situation.

6.7 Conclusions

It can be concluded from this part of the study that application of silence removal techniques to the outputs of the CoBliSS algorithm does not effectively moderate the adaptation and therefore fails to improve the overall performance for the filter lengths considered. The work did, however, raise some interesting issues that are worthy of consideration in the use of such

techniques : firstly, the importance of the appropriateness of the selection of the separation algorithm, which must take into account the temporal properties of the signals to be separated and the type and complexity of the filters to be resolved; and secondly, the selection of an assessment technique which is in the same domain (time or frequency) as that in which the updates are calculated — to ensure that an appropriate level of control over specific filter weights can be achieved, at the desired resolution or rate. This is particularly relevant in systems that process blocks of data in either domain. Finally, careful consideration must be given to the length of any such blocks of data being processed, independent of the length of the filter being sought, if this type of controlled adaptation is to be attempted. It is suggested that the length of the data blocks should be related to the characteristics of the data, such as the stationarity threshold for bursty data (*e.g.* speech).

6.8 Areas for further investigation

This research has identified several other avenues for investigation, which it was not possible to pursue within the timeframe of the thesis. These areas could further extend the results presented in this thesis, and would provide an interesting starting point for future project work :

- The effect of shaping the energy threshold for the different frequency bins when applying the silence removal technique in the frequency domain.
- The use of shorter block sizes or filter lengths to enhance the observation of the short-term spectral characteristics of the speech signals, and the necessary modification of the algorithms used to accommodate this.

6.9 Summary

In summary, this chapter investigated the application of the silence removal technique — used, as in the earlier chapters, as a means of non-stationarity reduction — to a decorrelation based blind separation and deconvolution algorithm, CoBliSS, in both the time domain and the frequency domain. No significant improvement was found in either case, and for the frequency domain variants, numerical problems were encountered. Likely reasons for these complications were identified, and potential solutions proposed.

Chapter 7
Summary of Conclusions and Future Work

Each chapter in this thesis is largely self contained, providing a more detailed discussion of the topics and experimental results contained in that chapter. The following provides an overview of the study, highlighting the contributions to knowledge made during the study, drawing conclusions from the research and identifying some ideas for possible directions of future work.

7.1 Review

The content of each of the preceding chapters is summarised here, before the principal conclusions of the investigation are emphasised.

Chapter 1 introduced the subject area and set out the aims of the study — namely to examine the effect of silence removal on the performance of blind separation of non-stationary sources, focusing on speech signals. The research was carried out using systems based on the information-maximisation framework proposed by Bell & Sejnowski [10] for instantaneously mixed signals, and on the CoBliSS algorithm by Schobben & Sommen [108, 123] for convolutively mixed signals.

Chapter 2 provided background information on different types of and approaches to signal processing, such as adaptive filtering and blind signal processing. Artificial Neural Networks, another architecture suitable for adaptive processing, were introduced and the learning strategies employed discussed. These were identified as being appropriate for use in solving blind signal processing problems such as blind signal separation and deconvolution.

Chapter 3 described the blind signal separation and deconvolution problems, looked at applications in which solutions to these problems could be useful and identified issues that must be considered in these solutions. A variety of approaches to solving the

ICA/BSS problem and the blind deconvolution problem were presented, covering both neural and non-neural techniques in both the time domain and frequency domain where appropriate.

Chapter 4 presented details of the experimental research undertaken on instantaneously mixed signals. Having identified speech signals as exhibiting the non-stationary characteristics known to detrimentally affect the performance of information-maximisation-based blind separation algorithms, results from experiments performed confirmed that a significant part of degradation in performance is due to deviation of the weights from their convergence trajectories during periods of silence of the source signals. As a result of initial experimentation, it was suggested that separation performance may be improved by eliminating these periods of silence, to yield a signal of more stationary variance, which would not result in so many or such serious deviations.

To test this suggestion, a series of tests were devised and carried out, pre-processing the source signals in question and assessing the resulting signals' state to identify periods of silence. Two methods of assessment were considered, strict thresholding and average energy thresholding. The average energy assessment was selected for use because of its better noise tolerance. The periods were characterised by both duration and threshold, which lead to a range of signals being generated with varying degrees of non-stationary variance. A metric based on the Co-efficient of Variation was devised to allow assessment of the degree of non-stationarity, and used to evaluate the effect of the pre-processing. This metric was also used as an index for the relative separation performance of a number of mixtures created from the pre-processed sources, using mixing matrices of varying complexity to test the range of applicability of the non-stationarity reduction technique. The effect of batch sizing used during the simulations was also considered.

The experimental approach enabled trends in the separation performance to be identified by observation of the graphed output. Statistical analyses using ANOVA facilitated comment on the separation performance by comparing the end points of the traces after a fixed number of iterations.

Chapter 5 extended this work, focusing it on the issues pertinent to the incorporation of the silence removal techniques into an on-line blind signal separation system. The system was developed in a novel way to conform to these requirements and reconfigured to continually assess its outputs and control the weight set update according to the

silence assessment. Comparisons, by visual inspection of graphs and statistical analysis, were made between the performance results of the systems using on-line and off-line processing respectively.

The effect of several additional methods on the performance of the on-line system were tested. These methods provided alternative update strategies to be employed when the output assessment indicated a period of silence. The results of these tests were compared to one another and to the results of an on-line permutation approach.

The general applicability of the work was then tested by comparing the performance of the network when processing different numbers of inputs. Alternative network architectures and processing strategies were also considered. The algorithms were modified to incorporate the same on-line silence assessment and update strategies as had been applied to the infomax algorithm, and their performances examined.

Chapter 6 investigated the application of these techniques to convolutively mixed speech signals. Various algorithms were considered, and the CoBliSS algorithm, by Schobben & Sommen [108, 123], was selected for use in the experimentation. Filters with simulated room characteristics and echoes at designated intervals were designed to create mixtures on which to test the performance of the algorithm. The effectiveness of the application of the silence removal techniques to both the time domain and frequency domain representations of the outputs was tested.

7.2 Conclusions

This thesis studied the effect of non-stationarity reduction techniques on the performance of blind signal separation and deconvolution systems. The research investigated the separation of both instantaneously mixed and convolutively mixed speech signals. For instantaneously mixed signals, the work was largely based around the information-maximisation blind signal separation system of Bell & Sejnowski [10]. Other systems considered were a natural gradient version of the infomax system, based on that proposed by Amari *et al.* [58] and the approaches of Jutten & Hérault [56], Matsuoka, Kawamoto & Ohya [51] and Barros & Ohnishi [53]. For the investigations on the convolutively mixed signals, the work focused on the use of the CoBliSS algorithm.

The use of silence removal was shown to reduce the degree of non-stationarity of a signal's

variance, which has a large effect on the convergence of the separating network's weight set, when that signal is one of the mixture sources. The experimentation showed that the off-line pre-processing of the source signals, prior to their instantaneous mixing, led to faster convergence of the separating network to a separating solution — particularly when using the average energy based assessment and when considering periods of longer durations. Results from early studies by Alphey *et al.* [127, 128] suggested that by reducing non-stationarity, silence removal may help to enhance separation. The more detailed studies developed in this thesis have shown that although the rate of separation was improved, the separation performance, as measured by the Amari metric, was not. Initial observation of the graphs suggested that non-stationarity reduction often appeared to give a small improvement in separation — however, statistical analyses of the results showed that no significant improvements were made.

The silence assessment techniques were subsequently incorporated into an on-line adaptive ANN-based separation system and used to control the update of its weights. The results achieved using the on-line approach were found to be similar to those of the off-line approach. Performance on mixtures of varying complexity, both in terms of the determinant of the mixing matrix and of the number of mixed signals, was investigated. For the unmodified algorithm, the results were as expected — separation performance being better for the more well-conditioned mixing matrices than the less well-conditioned. The performance of the modified algorithms was found to parallel that trend. Silence removal did not significantly improve the separation performance and at a threshold level of 10.00% significantly reduced the performance. Use of the alternative update strategies did not provide any additional benefit, and in several cases the separation performance was further reduced. The relative performance of other similarly modified BSS algorithms, some specifically designed for processing non-stationary signals, was also considered. Of the four algorithms considered, none were found to benefit from the use of the non-stationarity reduction techniques. The separation performance of the CoBliSS algorithm when applied to convolutively mixed signals was poor in comparison to that of the instantaneous separation experiments. Against this poor-performance baseline, the non-stationarity reduction techniques showed no significant improvement over the separation performance of the original algorithm.

These findings do not align with those of Nguyen Thi & Jutten [9] and Van Gerven & Van Compernolle [6–8], who considered similar approaches to control the update of the separating

system's weights. Both sets of authors reported significant improvements in the separation performance of the decorrelation-based systems that they used, when using energy-based assessments of the output signals to determine whether or not to adapt the weights. It would be difficult to definitively identify the specific factors that contributed to these differences in the reported effectiveness of such approaches, but they are likely to be linked to the complexity of the filters considered, and the conditional update strategies employed. In the experiments undertaken for this thesis, the filters used were far longer and more complex than those considered by the other researchers, and the strategies employed to determine when to modify the weights were based on different criteria, exploited differing levels of selectivity.

Non-stationarity in speech signals It is well known that speech signals exhibit non-stationary characteristics, principally a rapidly changing variance, over periods greater than about 20–25 ms. It has been shown that this particular type of non-stationarity leads to poor separation performance in information-maximisation-based blind separation systems — a fact confirmed by the work of this thesis.

Inter-word silences The periods of greatest variance non-stationarity are those that occur as the speech signal moves from a region of active speech into one of inactivity, or relative silence, such as happens between spoken words. It was shown in Section 4.2 that this is also where the weights of the infomax-based separation system suffer the largest disturbance from their convergence trajectories. This disturbance was found to lead to poor performance.

Earlier work by Nguyen Thi & Jutten [9] and Van Gerven & Van Compernolle [6–8] on convolutive mixtures found similar results — that updating the network only when there was sufficient energy present gave better separation performance.

Non-stationarity reduction Eliminating such periods of silence resulted in a much smoother variance envelope over the source signals, and a reduction in the corresponding non-stationarity assessment metric for the signal. The use of simple, energy-based silence identification and removal techniques was shown to be effective at this task, and resulted in faster convergence to a separating solution in tests carried out on mixed speech signals created from such processed sources. The duration parameter was found to have no significant interaction in many of the experiments considered.

Batch sizing When using batch processing to reduce the computational load, the size of the

batches used was found to have a significant effect on the separation performance of the more well-conditioned mixtures. In these cases, longer batch sizes of 200 samples were found to give better results than shorter batches of 50 samples.

On-line performance Since the off-line pre-processing of source signals to reduce their non-stationarity could not be used in an on-line blind separation system, a modified approach was devised that controlled the update of the system's weight set based on an assessment of its outputs — the estimated sources. This new system exhibited faster convergence than the original infomax system under certain conditions, but did not offer any significant improvement in separation performance.

Additional strategies that provided alternative updates during periods of silence were also investigated. These approaches, namely the average delta, buffered inputs and variable learning rate approaches, were not found to offer any significant improvement over the basic on-line assessment, and in some cases reduced the separation performance.

Instantaneous mixture complexity The results from the new system devised in this research mirror the trend in separation performance of the original algorithm on mixtures of varying complexity — performing better on well-conditioned matrices. The spread of results over different silence removal parameters was wider for the more well-conditioned matrices.

Choice of algorithm The performance of several other separating algorithms, each modified to incorporate the on-line silence assessment modifications and alternative update strategies, was compared. None were found to produce significantly better results after the modification, but their relative performances indicated that the natural gradient and the pre-filtering versions of the infomax algorithm gave the fastest convergence and best separating performance. Although the pre-filtering algorithm of Barros & Ohnishi gave considerably faster convergence, it was found to be unstable under a variety of conditions. The Hérault-Jutten network was found to converge consistently to a non-separating solution for some of the mixtures considered, and the Matsuoka, Kawamoto & Ohya network showed noisy but good convergence on the more well-conditioned matrices.

Convolutive mixing Filters with typical room impulse characteristics and strong echoes at key intervals were designed, and used to create acoustic mixes of the source signals used

previously. Separation and deconvolution using the CoBliSS algorithm was found to give poor results, compared to those achieved in the earlier experiments.

Time domain versus frequency domain The on-line silence assessment methods were incorporated into the algorithm, and used to selectively control the update of the weight set according to the assessment of either the time domain or frequency domain representation of the outputs. The results from the time domain assessment showed no significant improvement over those of the original algorithm. The frequency domain assessment experiments fared particularly poorly, and many did not run to completion.

Conditional adaptation The results of this thesis differ from the earlier findings of Van Gerven & Van Compernolle [6–8] and those of Nguyen Thi & Jutten [9] who reported beneficial results from the application of conditional update control. However, as discussed previously the filters used in their experiments were considerably shorter and less complex than those used in the investigations carried out in this thesis, and the approaches taken to the conditional update differed in their focus and their application.

This study has extended the knowledge in the field of blind signal separation. The conclusions of this research, in addressing the aims of the thesis, show that the silence-removal-based non-stationarity reduction techniques do not offer an effective means of improving the separation performance of infomax-based systems or the CoBliSS system tested. The results of this thesis, therefore, do not support the original hypothesis that non-stationarity reduction of speech signals by silence removal would benefit the performance of the blind separation systems tested in this study.

In many instances the separation performance, following the non-stationarity reduction, was poorer than that of the unmodified algorithms considered. It seems that some periods of silence within the source signals may aid separation, possibly by causing a degree of non-stationarity in the surface of the objective function that assists the system in escaping from local extrema of the solution space. In the instantaneous mixing cases considered, silence removal did reduce the degree of non-stationarity of the signals' variances and helped to improve the rate of convergence of the separating systems' weight sets. It may therefore be useful where this rate of convergence is important.

There was no significant improvement in the separating and deconvolving performance of the modified CoBliSS algorithm when the outputs were assessed in the time domain, and

the frequency domain assessment resulted in convergence difficulties rendering the approach unreliable. Therefore, the use of silence removal cannot be recommended as a means of improving infomax-based systems in blind signal separation, nor for improving the separation of convolutive mixtures.

7.3 Future work

In addition to the work presented above, this study has identified other lines of research which could not be addressed in the time available. Suggestions of areas worthy of further study to extend this interesting field of research are listed below. Some areas are of particular relevance to lines of research undertaken in this study, while others are more general in nature and may be of interest to the wider BSS / ICA research community, or to application developers dealing with these types of problems.

- Investigation of the differences in the separation performance of the algorithms considered due to temporal correlations should be carried out. The impact of silence removal or other non-stationarity reduction techniques on the separation performance on both of these classes of signals should be assessed for both instantaneous and convolutive mixtures.

- Comparisons between the work documented in this study and that concerned with non-stationary mixing environments should be made. This would facilitate a "generalised non-stationarity" approach to be developed, or at least guidelines to be proposed for dealing with common features of the presence of non-stationarity.

Appendix A
Published Papers

This appendix contains re-prints of the papers [127,128] published externally during the course of this research :

127 M. J. T. Alphey, D. I. Laurenson and A. F. Murray, "The effect of signal non-stationarity on the performance of information-maximisation-based blind separation" in *Neural Networks for Signal Processing VIII — Proceedings of the 1998 IEEE Workshop (NNSP98)*, pp. 113–22, September 1998.

128 M. J. T. Alphey, D. I. Laurenson and A. F. Murray, "Improvements in the on-line performance of information-maximisation-based blind signal separation" in *Proceedings of the First International Workshop on Independent Component Analysis and Blind Signal Separation (ICA'99)*, pp. 49–54, January 1999.

The results published in the papers are from individual experiments performed in the early stages of the study. The corresponding results presented in the thesis, in Chapters 4 and 5, are from more extensive studies based on the earlier work and using results averaged over different starting points and times.

To accommodate the papers in this thesis, their pages have been rescaled to reduce them slightly in size.

A third paper [129], which formed the basis of paper [127] above, was published in an internal departmental journal, but is not reproduced here :

129 M. J. T. Alphey, D. I. Laurenson and A. F. Murray, "The effect of signal non-stationarity on the performance of information-maximisation-based blind separation" in *PhDEE — The Postgraduate Journal of the Department of Electronics and Electrical Engineering*, Issue 4, pp. 59–65, April 1998.

A.1 1998 IEEE Workshop on Neural Networks for Signal Processing VIII (NNSP98)

The Effect of Signal Non-Stationarity on the Performance of Information-Maximisation-Based Blind Separation

M. J. T. Alphey (*Marcus.Alphey@ee.ed.ac.uk*)
D. I. Laurenson (*Dave.Laurenson@ee.ed.ac.uk*)
A. F. Murray (*Alan.Murray@ee.ed.ac.uk*)

Department of Electronics & Electrical Engineering,
The University of Edinburgh,
The King's Buildings, Mayfield Road, EDINBURGH, EH9 3JL.

Tel : (0131) 650 5565 *Fax : (0131) 650 6554*

Abstract

This paper examines the performance of a blind signal separation system that uses information maximisation techniques, when applied to signals with non-stationary characteristics. It assesses the effectiveness of different methods of removing or reducing the degree of non-stationarity of the source signals, in terms of the level of separation achieved after a fixed training period.

1 Introduction

Blind signal separation is an area of research that is currently benefitting from a great deal of interest in fields including signal processing and neural networks, to name but two. The basic problem is to separate m linearly mixed signals back into their n mutually independent sources, without any other *a priori* knowledge about those sources, nor their mixing.

Much of the current research into blind signal separation is based on work by Bell and Sejnowski[1], which applied elements of information theory to a neural network based solution to the problem. Their paper, like many others, makes the assumption that the signals to be separated have stationary characteristics. Here we investigate ways of improving the performance of such systems when this is not the case.

2 Background

The blind signal separation (BSS) problem is illustrated below in Figure 1. The existence of any echoes is ignored, and each sensor receives a different (linear) mixture of some or all of the sources, at any instant in time. The outputs of these sensors form

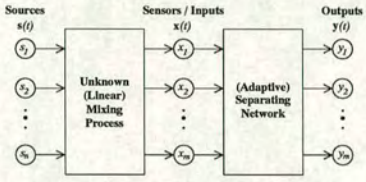


Figure 1: An illustration of the blind signal separation problem

the inputs to the separating system, which is configured to have the same number of outputs as it does inputs. However, the number of sources is unknown.

As in many of the papers in this area, we will assume that a separating solution exists, and will focus our attention on cases where $n = m = 2$.

2.1 Information Maximisation

Bell and Sejnowski's work focussed on a neural network based system, and took an information-theoretic view of the signal separation. Much of this was based on Linsker's work on information maximisation - his Infomax Principle[2]. Linsker's observation was that, given a fairly basic learning rule, each layer of an unsupervised network could adapt to minimise the mutual information between each of its individual outputs, thereby reducing the overall redundancy of the output layer.

When applied to the problem of blind signal separation, this has the desired effect of reducing the level of mixing in the outputs up to the point where they contain no redundant information about each other, and hence are fully separated.

Using just a single layer network (see Figure 2), Bell and Sejnowski reported very good separation results from mixtures of as many as ten original sources. However, in their discussion of their experimental procedure they note that permutation of the time index of the samples prior to mixing was necessary to "ensure that the input ensemble was stationary in time". In a more recent paper, Barros and Ohnishi[3] note that this eliminates any possibility of using this technique for online separation. They propose a modification to the setup that allows non-stationary signals to be separated in an online fashion as well, by pre-filtering the inputs, since it is desirable that both stationary and non-stationary signals can be dealt with. The reason that the

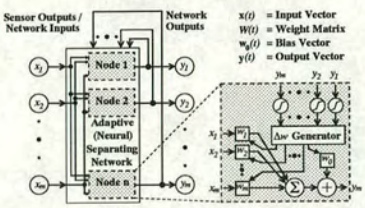


Figure 2: Single layer separating network

information maximisation technique does not work so well on non-stationary signals is that it is a relative gradient algorithm, and as such, will be affected by changes in the mean of the signal.

2.2 Performance Measures

In order to allow a comparison of the performance of the separation system under different conditions, some method of quantifying the degree of separation was required. The method adopted was that proposed by Barros and Ohnishi[3], which uses the ratio of the absolute values to the sum of those values for each column of the matrix $\mathbf{C}$, where $\mathbf{C} = \mathbf{AW}$. These separation values, p, will all be in the range $0.5 < p < 1.0$, with $p = 1.0$ indicating complete separation. Expressed mathematically, for an i by j matrix $\mathbf{C}$, the ratio is :

$$p_j = \max\left\{\frac{|c_{i,j}|}{\sum_j |c_{i,j}|}\right\}|_i \quad \forall i, j \tag{1}$$

with the provision that each i must map to a unique j, and *vice-versa*.

3 The Effect of Signal Non-Stationarity on the Performance of Information-Maximisation-Based Blind Separation

The objective of this part of the investigation was to assess the degree to which non-stationarity of the source signals affected the separation performance of neural network systems using information-maximisation-based update rules. A parameterised software model of the system was constructed to enable tests to be carried out, and was configured to implement Bell and Sejnowski's information maximisation arrangement with two inputs and two outputs, for each of the experiments. The network weights and biases were randomly set between -1.0 and 1.0 before training commenced.

A number of tests were run with various input signals, and the weight sets and bias vectors of the network were recorded throughout the duration of the tests, as it was these that were of interest here more than the outputs. In fact, the outputs were only used to confirm results indicated by the other data.

3.1 Reduced Silence Source Data

Since the focal point of this investigation involved the non-stationarity of signals, the source data to be used were all selected so that they exhibited this desired characteristic. The data sets from which these source signals were drawn were speech and music data, sampled at 8 kHz with 12 bit resolution. Mixing of the source signals, to create the network inputs, was achieved by multiplying them with an artificial mixing matrix. This meant that the required separating matrix could be calculated directly, to allow analysis of the experimental results.

Some of the source signals were preprocessed prior to mixing, to create a range of input signals with varying degrees of non-stationarity. Since one of the main sources of non-stationarity in a speech signal is the periods of silence between the sounds, a simple way of varying the level stationarity exhibited by a signal is to remove these intervals. In doing so, there were a number of choices to be made, all related to isolating the periods of silence from the rest of the signal. The three main choices were :

- the method by which periods of silence to be examined were to be judged
- the duration of the period of silence to be removed
- the level below which the signal would be judged silent

Varying the parameters of the silence-removal process resulted in mixed inputs with different lengths, but to allow for a fair comparison between them, all of the experiments were run for 400 000 iterations. This value was chosen as it gave five passes through the original (unprocessed, and therefore longest) input signals, normally enough to permit most of the separation to occur.

Others sources had their time indices permuted, again prior to mixing, as Bell and Sejnowski had done, to remove all traces of non-stationarity.

3.1.1 Method of Identification

Two methods were considered for classifying which regions of the source signals were to be deemed silent. The first was a straightforward comparison of the sample data against a threshold level, for a number of consecutive data values.

The second method involved performing an energy equivalent comparison, which used the average squared value of the consecutive samples.

Source signals were generated using both methods, for a fixed threshold level, and over a range of silence durations. The performance measures for each individual run were then calculated and averaged. Figure 3 shows the comparison of these averages for the two methods, for each pair of corresponding signals. The energy equivalent

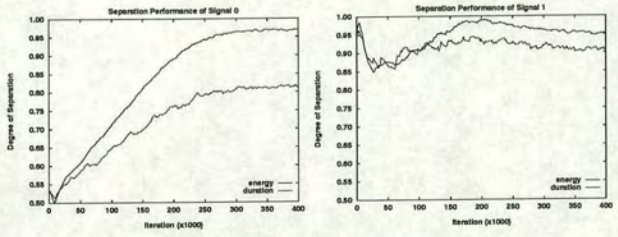


Figure 3: Performance with different methods of silence identification

method reaches a higher average level of separation, and reaches this level more quickly than the other method. This is to be expected, as the former provides a more accurate representation of the level of the signal, as it is more noise tolerant.

3.1.2 Duration

Altering the period over which the signal is to be assessed for silence, can have a profound effect on the output, especially for rapidly changing signals such as speech. A range of signals were generated having had periods of silence removed from them. These periods were of duration of as few as 2 and up to as many as 8000 samples (*i.e.* one second at the given sample rate).

Shown in Figure 4 are the results of a comparison between signals which have had silences of 2, 500 and 8000 samples removed, respectively. Perhaps surprisingly, it is the signal with the longest silences removed that gave the fastest and highest average separation performance. Intuitively, it would seem that removing smaller periods of silence would result in a much more stationary signal, as these rapidly changing sections would no longer have the sudden drops to zero (or near zero) at the end of each burst of speech, and it seems likely that there should be a larger number of shorter sequences that satisfy the silence criteria than longer ones. However, it is the loss of these sudden changes that actually causes the decrease in performance illustrated in the graphs, as they constitute important information about the speech itself. By removing them, this information is lost to the adapting network, and so it takes longer to reach the correct separating weight set. The longer delays, by

comparison, tend to correctly remove silences between separate bursts of speech, whilst leaving the 'silences' within the speech intact. Obviously, there will be a critical duration which separates intra- from inter-signal silences, but this may well vary from signal to signal.

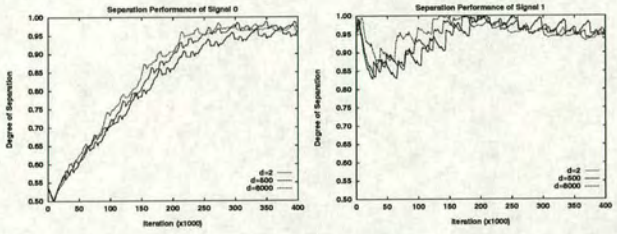


Figure 4: Performance with varying durations of silence

3.1.3 Threshold Level

Altering the threshold below which a signal is judged to be silent has a very obvious effect on the output - the higher the threshold, the more energy the signal must contain over the window of interest. It can be seen in Figure 5 that a higher threshold leads a higher level of separation, and that this separation occurs more quickly than for lower thresholds. (l is a parameter used in the energy estimation, lower values of l leading to higher threshold levels.) However, with too high a threshold (see the graph for $l = 4.0$) too much of the original signal is lost, and although the apparent separation performance is very good, the output is unacceptable. A balance between the desired quality of the output and the speed or degree of separation must be found, but again, this is likely to depend on the signals used.

The peak and decay in the separation performance graphs of signal 1, most clear for $l = 4.0$, is due to the competitive nature of the algorithm. In order to improve on the separation of signal 0, and the overall separation performance of the system, signal 1 must endure this reduction in separation quality, despite it having reached near full separation for a short while.

3.1.4 Permutation

The final comparison carried out in this set of experiments was on the effect of permuting the time index. Figure 6 shows a comparison of the separation performances for the original and permuted sources.

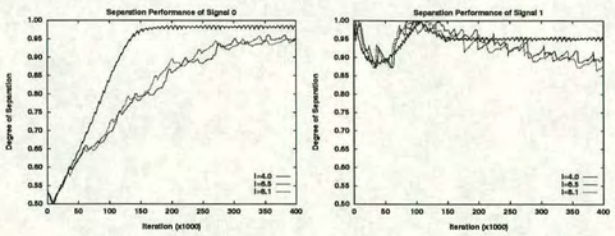


Figure 5: Performance with varying threshold levels

It can be seen that the graphs converge to approximately the same levels, and that these are lower than those achieved by the some of signals preprocessed using the silence removal methods. The smooth shape of the graph for the permuted sources is due to the removal of any time dependent correlations within the signal. These are what gives rise to the regularly repeating patterns visible in the other graphs.

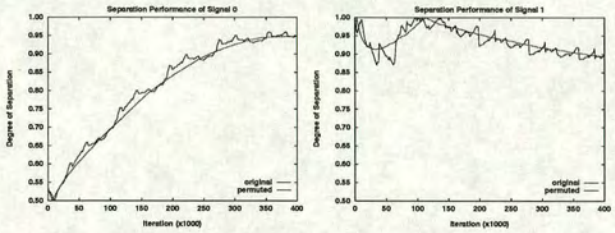


Figure 6: Performance with and without permutation

4 Non-Stationarity Reduction in an Online Signal Separating System

In order to make these techniques amenable to online processing, the silence removal operations required to be incorporated into the information maximisation update rule for the network's weight set. However, since the network receives as its inputs only the mixtures of the source signals, the best that can be done is to remove periods of silence from these instead. Doing so has less effect on the non-stationarity of the mixture than removing periods of silence from the source signals directly, as fewer periods of silence are likely to exist in the input signals. Attempts were made to improve on this situation by applying the energy estimation technique to the network's current estimate of the outputs, in evaluating when a source signal was silent.

Having made the above changes changes to the algorithm, to produce what will from now on be referred to as the "modified algorithm", the experiments varying the duration and threshold values of the silence identification criteria were rerun. This time, the original signals were used in their entirety, mixed together without any preprocessing. The d (duration) and l (level) parameters for the silence removal algorithm were simply passed to the software model, thereby eliminating the need to carry out any off-line preprocessing.

The experiment involving the permutation of the source signals was not repeated, as it would not be practical in a real-world situation, where the length of the entire signal could not be available in advance. Hence it would be impossible to generate a true permutation over the entire signal. A partial permutation over a windowed area of the signal could be carried out, but stepping this through the signal may not be as effective at removing all traces of non-stationarity as a full permutation. This would be particularly true in situations when the duration of the periods of silence tended to be large, relative to the windows sizes used.

4.1 Results from the Modified Information Maximisation Algorithm

The same range of experiments were run as before for varying durations and threshold levels. (Full sets of results omitted due to space restrictions.) The results of the experiments using the modified information maximisation algorithm agree closely with those obtained from preprocessing the source signals, as illustrated by Figure 7 for the example case of a silence duration of $d = 20$ and a threshold level of $l = 6.5$. In all but the case for $d = 20$, $l = 4.0$, the overall results are at least as good as those achieved by the original algorithm, although the individual performance of each signal may not necessarily be better for any given case (again, as illustrated in Figure 7).

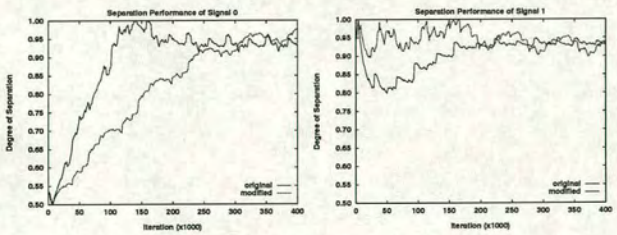


Figure 7: Comparison of algorithm performance with $d = 20$, $l = 6.5$

5 Analysis of Results

The preprocessing of the source signals used in these tests results in notably better separating performance than for the unprocessed signals, when used with the original information maximisation algorithm. However, this does not provide an ideal solution to the problem of dealing with non-stationary signals, as it depends on the source signals being available for this preprocessing - something that is unlikely in practice. The modified algorithm on the other hand provides the benefits of these processing techniques in a manner that places no more constraints on the problem than the original information maximisation solution itself.

From the graphs of the comparison of outputs it can be seen that the modified algorithm may not converge as quickly as the original. This is due to the algorithm having to process data that would otherwise have been removed by the preprocessing stage. However, the level of separation reached after the 400 000 iterations is approximately the same. Furthermore, this is quite a high level of separation, although a trace of the other signal is still present.

Both techniques have outperformed the results obtained when running the original information-maximisation-based algorithm on unprocessed source mixtures, but the online algorithm's performance is more significant as this will be of more practical use in real-world situations. The use of these techniques are not limited to speech nor music mixtures, but as was noted earlier, careful selection of silence identification parameters (d and l) may be required for some specific cases.

Other results indicated that further simulation for longer time periods yielded little improvement in the final level of separation attained. After the initial 400 000 iterations (50 seconds) shown, the outputs had settled to a reasonably stable state, and whilst minor convergence was noted beyond this point, in most cases it was largely already completed.

6 Conclusions

From the experimental results it is clear that signals with non-stationary characteristics can be separated by neural network systems employing online information-maximisation-based learning rules. The more stationary the signals' characteristics, the faster the signals are separated, and after a fixed number of iterations, the higher the degree of separation achieved.

Silence is a major contributing factor to the degree of non-stationarity in speech signals, and it is possible to preprocess such signals to improve the separation performance in a number of ways, or to incorporate these techniques into an algorithm to be used online. The use of an energy estimation method for identifying periods of silence proved more accurate than strict thresholding. The duration of the silences to be removed should be set so that inter-speech silences are removed whilst retaining

intra-speech silences. The threshold level of the energy estimating silence identification algorithm should be set so that sufficient signal remains for the network to adapt on, having discarded the unwanted silence and noise. Off-line permutation of source signals also leads to a good separation performance, but would not, in general, be practical in online algorithms.

The incorporation of silence removal techniques into online information maximisation based neural network update rules is at least as effective as the off-line preprocessing of sources signals, in terms of the degree of separation achieved. In terms of convergence time, the original information maximisation algorithm performs slightly better, when used with the preprocessed sources. However, the modified algorithm is of more potential benefit as the preprocessing option may not always be practical or even possible. Both of these silence removal techniques offer notable improvement over the standard separation case of the original algorithm and the unprocessed source signals.

7 Acknowledgements

This work was sponsored by GEC Marconi Avionics, and the data used was provided by BT Laboratories, Martlesham Heath, Ipswich.

References

[1] A. J. Bell and T. J. Sejnowski, "An information-maximisation approach to blind separation and blind deconvolution", *Neural Computation*, vol. 7, no. 6, pp. 1129–59, November 1995.

[2] R. Linsker. An application of the principle of maximum information preservation to linear systems. In D. S. Touretzky, editor, *Advances in neural information processing systems*, vol. I, pages 186–94. Morgan-Kaufmann Publishers, Inc., 2929 Campus Drive, San Mateo, CA 94403, 1st ed., 1989.

[3] A. K. Barros and N. Ohnishi, "Pre-filtering non-stationary signals to improve blind source separation", presented at *Proceedings of the 13th International Conference On Digital Signal Processing*.

A.2 First International Workshop on Independent Component Analysis and Blind Signal Separation (ICA'99)

IMPROVEMENTS IN THE ON-LINE PERFORMANCE OF INFORMATION-MAXIMISATION-BASED BLIND SIGNAL SEPARATION

M. J. T. Alphey, D. I. Laurenson and A. F. Murray

Department of Electronics & Electrical Engineering, The University of Edinburgh,
The King's Buildings, Mayfield Road, EDINBURGH, EH9 3JL, Scotland, UK.
Email : {mjta, dil, afm}@ee.ed.ac.uk

ABSTRACT

This paper investigates the effect on separating performance of certain modifications to the update rules of an information-maximisation-based blind separation algorithm. These modifications are intended to improve the rate at which signals can be separated by an on-line system, with little additional computational cost.

1. INTRODUCTION

Blind signal separation has recently attracted a great deal of attention from both the neural network and signal processing communities. As interest in this paradigm grows, the range of problems to which it is applied also expands. Various generalisations, extensions and techniques now exist which allow adaptive separation of signals [1] as well as separation of those with real world [2] and non-stationary characteristics [3, 4]. Of the various techniques that can be used, the most popular and powerful are Bell & Sejnowski's Information Maximisation-based approach [5], and Independent Component Analysis (ICA), as formalised by Comon[6]. The former avoids some of the necessary approximations of the latter, but does not necessarily lead to a statistically independent set of outputs, as shown by [7]. Amari, Cichocki and Yang[8] have also suggested improvements to this method, notably the use of natural gradient based rules. In this paper, simple modifications are presented that yield noteworthy improvements in the rate of separation.

2. BACKGROUND

The basic aim of blind signal separation is to separate out n unknown, individual, independent source signals from m observed, linear mixtures without any *a priori* knowledge about either the sources or their mixing. For the purposes of this paper, it is assumed that a separating solution exists, and that $n = m = 2$ for simplicity.

2.1. Information Maximisation

Bell & Sejnowski's work [5] used a single layer neural network (see Figure 1) and approached the separation problem from an information theoretic point of view, based on Linsker's principle of maximum information preservation, or *infomax* principle [9].

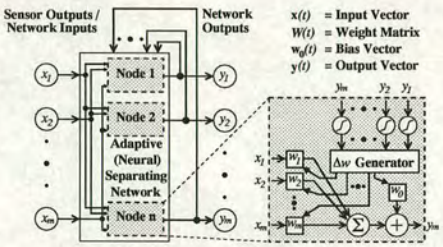


Figure 1: Single layer separating network

Linsker's observation was that given a fairly basic learning rule, each layer of a unsupervised network could adapt to maximise the information transfer between its inputs and its outputs, whilst at the same time minimising the mutual information between those outputs. In terms of the signal separation problem, this minimisation of mutual information has the effect of reducing the redundancy (and hence the level of mixing) at the outputs, thereby separating the signals.

The neural perspective on this process is that the network's weight set $\mathbf{W}$ 'learns', or converges to, the values of the matrix required to separate the mixed signals. This will be a (potentially) scaled and permuted version of the inverse of the mixing matrix $\mathbf{A}$. The

gradient ascent learning rules of Bell & Sejnowski's algorithm that enable this convergence are based on the entropy of the system, relative to the network's weight set $\mathbf{W}$ and bias vector $\mathbf{w_0}$. They generate update terms which are added to the current value of $\mathbf{W}$ or $\mathbf{w_0}$, iteratively moving the network towards its desired solution.

2.2. Potential problems

While all the sources are providing useful information, the algorithm works well, since all the outputs compete with one another - an increase in the value of the weights contributing to one resulting in a decrease in the weights related to others. However, if at some point one or more of the sources fall silent, the weight set and bias updates (or *deltas*) generated by the algorithm may no longer lead in the direction of the original separating solution. This is because the algorithm will continue to attempt to fulfill its learning rules, despite the input conditions having changed. The system is lacking information from one of the signals, and the deltas generated cannot take account of the contribution that this signal would have made to the separating solution.

2.3. Proposed solutions

It is possible to calculate an estimate of the energy level of the outputs, and from this to determine whether or not each meets some predefined threshold below which they can be considered to be silent, or to contain insufficient useful information. The update rules can be modified to take account of this assessment, so that the weights and bias values depending on the output in question should not be updated with the standard delta.

The alternative update strategies presented below make use of this energy assessment and replace deltas generated from the current inputs and outputs with updates based on a buffered set of successive input and output values. None of the adaptations investigated are computationally expensive, and since they use finite buffer sizes, they are all amenable to use in on-line systems.

2.4. Performance measures

The performance of the various separating algorithms was assessed by the degree to which the network weight set $\mathbf{W}$ matched the values of the inverse of the mixing matrix $\mathbf{A}$, allowing for scaling and permutation, at the end of the fixed-length simulation. This can be calculated by multiplying together the weight set $\mathbf{W}$ and the mixing matrix $\mathbf{A}$, which should yield a scaled, permuted approximation of the identity matrix $\mathbf{I}$. Using the relative values of the maximum value of each column to the sum of values for each column of this matrix, as proposed by Barros and Ohnishi [3] gives the desired figures. Mathematically, this can be formulated as :

$$p_j = \max\left\{ \frac{|c_{i,j}|}{\sum_j |c_{i,j}|} \right\}|_i \quad \forall i,j \tag{1}$$

where p_j is the separation performance of signal j, $\mathbf{C} = \mathbf{AW}$ and each i maps to a unique j.

3. ASSESSMENT OF SEPARATING PERFORMANCE

The investigation into the effect on the separating performance of the modified algorithms was carried out by means of computer simulations. A software model implementing an on-line information-maximisation-based blind separation algorithm (previously used to investigate the effect on on-line separation performance of non-stationary signals [4]) was configured to have two inputs and two outputs. A pair of speech signals (both sampled with 12 bit resolution at 8 kHz) were artificially mixed, using a known mixing matrix, to provide the inputs. For each of the different learning rules proposed, a range of simulations were run using the same pair of mixed signals in each experiment to allow a fair comparison of performance as the simulations progressed. The duration of all of the tests was set to 400 000 iterations, equal to five passes through the input data, as most of the weight set convergence had occurred by this point, even in the original algorithm. An initial set of weights and biases for the network was randomly selected from the range $[-1.0 : +1.0]$, and recorded so that each simulation could be started from the same point, again to allow comparison of the results. During the simulations, the values of the network's weight set and biases were recorded at each batch update (every 1000 iterations) and later used along with the known mixing matrix to generate the separation performance graphs.

All of the modifications to the infomax algorithm examined here require no special pre- nor post-processing, nor are they computationally intensive, and hence are suitable for use in on-line systems. This makes them of particular interest as one of the ultimate goals of this research is to develop a usable on-line separating system, for applications related to teleconferencing. It is only the calculation of the performance index that is done off line, and this was only done to quantify the success of the various techniques.

3.1. Threshold limiting

The first modification to the update rules buffers each of the outputs and estimates the energy levels over the specified number of buffered values, d. Should this estimate drop below a predetermined level l (the values shown relating to the information content of the data) for any of the outputs, updates on all weights or bias terms affected by that particular output signal are set to zero - *i.e.* no update is carried out for those value for that iteration. The weights and bias terms whose update is not affected by this signal adapt as normal. In this way, only updates based on information that will lead towards the desired separating solution should be generated.

A spread of results was generated using a range of durations and thresholds levels to determine when the output signals contained no useful information. A comparison of the separating performance of the original infomax algorithm and the best of these thresholding results (with a duration $d = 500$ and a threshold level of $l = 6.5$) is shown in Figure 2, along with the results of the next experiment.

As can be seen, the separation performance of the thresholding algorithm matches fairly closely that of the original infomax algorithm for (what was arbitrarily chosen to be) the first of the outputs, but shows nearly a 10% improvement on the final separation level of the second signal.

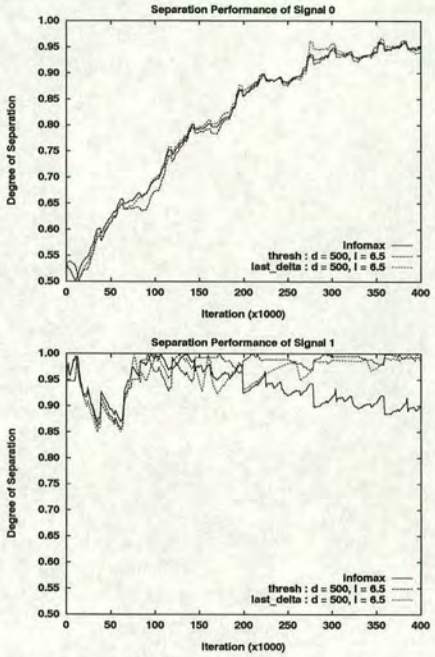


Figure 2: Comparison of infomax, thresholded and last delta performances

3.2. Last Delta

The next modification was an extension of the first result. Having determined the optimal duration and threshold upon which to base the decision of whether or not to update the network using the current inputs, the next step was to investigate if it was possible to replace the original update with one generated from some other set of values. Since the update generated is based on the gradient ascent rule, an obvious approach to try when no update is available is to keep going in the direction provided by the previous update.

At best, this should result in the algorithm achieving the ideal updates, although at worst, it could result in updates in completely the wrong direction, if the last 'good' update happened to have been on the brink of a turning point in the solution space. This technique may therefore, in certain circumstances, actually lead to worse results than doing nothing at all, but this will all depend on the signals present, and on the mixing environment.

The results shown for this experiment make use of the thresholding parameters determined by the previous tests, although other values were substituted to see if this in any way affected the results. From the graphs of Figure 2 it is clear that there is only marginal improvement in the separation performance of both signals, compared to the thresholding results, and that the second signal shows a little more instability earlier on.

3.3. Average Delta

The "last delta" rule can be viewed as a specific case of an "average delta" rule, which substitutes an update calculated from the average of the last a updates, where $a = 1$ for "last delta". Increasing the length of the buffer over which the average is calculated should improve the performance of the algorithm.

The reasoning behind this is that if the updates are correctly tending towards the optimal solution, using the gradient ascent rule, then their average should also progress the weight set and bias in that direction.

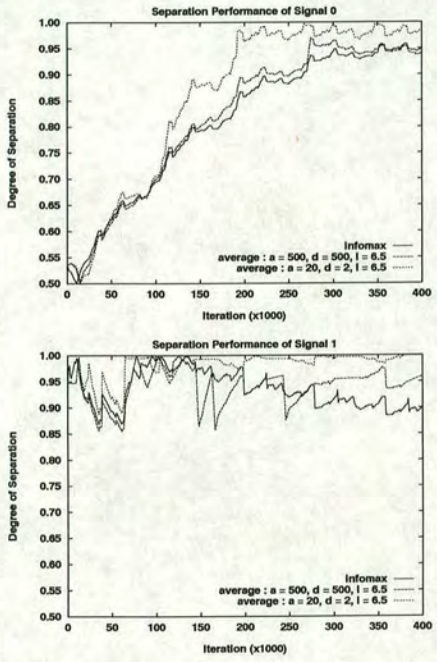


Figure 3: Comparison of the infomax and average delta performances, with varying parameters

Furthermore, should the prior sequence of updates be varying in the sign of one or more of its components, then this scheme would produce a much smaller delta for those particular components, whilst allowing ascent in the other directions to continue at a rate determined to the average. In this way, any maxima encountered by the algorithm in its ascent affects its performance only in the dimensions of those particular maxima.

Figure 3 shows a comparison between the infomax results, the best of the results obtained using averaging along with the thresholding limits determined by the previous experiments, and the best results obtained by running simulations with a range of buffer sizes (a), and varying threshold parameters. There is a clear improvement when a smaller a is used with shorter threshold durations.

3.4. Buffered Inputs

The final experiment presented here uses a slightly different approach. Whilst the outputs are still buffered and used to assess when an alternative to the normal update should be sought, this new update is now determined by randomly selecting an input from a buffer of the b best inputs seen so far. Hence this buffer is continually updated such that when the estimated energy of the b successive, most recent inputs exceeds that of the currently stored buffer, the buffer is replaced. In this way, the buffer always contains a set of data known to yield a useful output signal. Thus, when the current input fails to do so, it can simply be replaced by one that will.

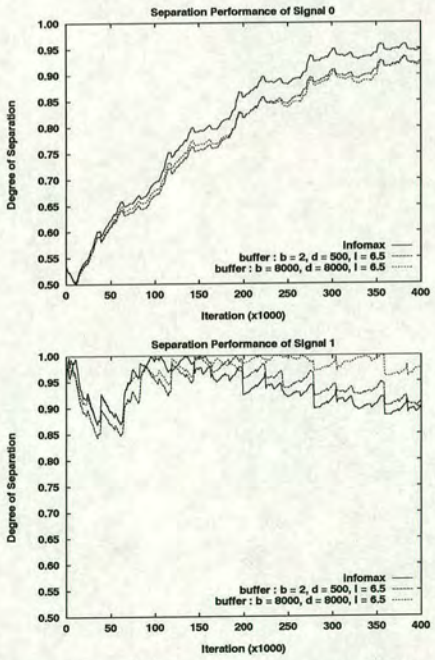


Figure 4: Comparison of the infomax and buffered input performances, with varying parameters

The results in Figure 4 show that using buffering along with the threshold values determined in the first experiment (Section 3.1) leads to an improvement in overall separating performance, and that even better

results are achievable when using larger buffer sizes and thresholding parameters.

Performance here might be expected to be better than that of the averaging results, given that the buffer should be full of useful data at all times, but as the graphs indicate, this is not the case. The reason for this is that while the buffer as a whole contains more information, the individual inputs at any one point may not necessarily all contain useful information. This is due to the way that the energy is estimated over the buffered outputs. Hence it is possible (and as happens to be the case for the examples given, as indicated by the results in Section 3.5) for the buffer to contain a high-enough proportion of low-information inputs, that randomly picking from it can lead to worse performance than from the average the a most recent updates. In these experiments, this is likely to be due to the non-stationary nature of the mixed speech signals used as the network inputs.

3.5. Comparison of results

Figure 5 shows a comparison of the best results obtained from each of the different techniques. As can be clearly seen from the graphs, the best performance is attained by using the average of the last 20 updates, when run with a very short buffer for assessing the signal energy. The worst performance is that of the method using buffered input to replace the current input, although this is still better than the performance of the original algorithm, *i.e.* of doing nothing at all. The thresholding technique results, whilst not as good as the averaging results, are still a noteworthy improvement on the those of the plain infomax algorithm, and in these tests outperformed the buffered input solution.

The final separating results are presented here as Table 1, although due to the fact that the buffered solution picks a (pseudo-)random sample to use for its input, the exact value that it attains is subject to some variation, whereas the others are all determinable.

Method	Total Separation
infomax	0.92414
thresholding	0.97228
average	0.99176
buffered	0.94771

Table 1: Overall Separation Performance (%)

Ongoing work is investigating the extension of these techniques to networks with more than two inputs and outputs, and to signals with different characteristics.

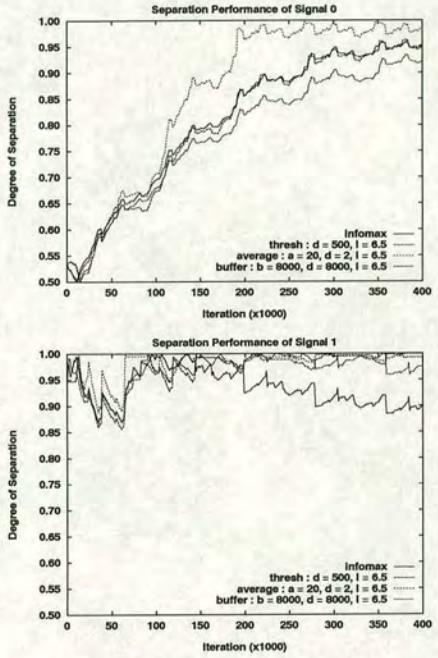


Figure 5: Comparison of the best of each of the investigated techniques

4. CONCLUSIONS

This paper puts forward a number of modifications to the information-maximisation-based blind separation algorithm, intended to yield improved on-line separation performance. Each of the methods is shown, by means of computer simulations, to be capable of producing faster separation results than the original infomax algorithm, at very low additional computational cost.

The plain thresholding technique, which does not update weight or bias components whose output signal does not meet some minimum energy level, offers a clear improvement over the original infomax algorithm. A solution that instead uses buffered input values as replacements for the current input when this fails to meet the necessary criteria, performed poorly by comparison. Finally, the best performance was achieved by replacing the update by one based on the average of a number of the previous updates.

5. ACKNOWLEDGEMENTS

This work was sponsored by GEC Marconi Avionics, and the data used was provided by BT Laboratories, Martlesham Heath, Ipswich, UK.

6. REFERENCES

[1] S. Choi and A. Cichocki, "Adaptive blind separation of speech signals: Cocktail party problem", presented at *International Confererence on Speech Processing (ICSP'97)*, pages 617–22, 1997.

[2] T.-W. Lee, A. J. Bell, and R. Orglmeister, "Blind source separation of real world signals", presented at *1997 IEEE International Conference on Neural Networks (ICNN'97)*, vol. 4, pages 2129–34, June 1997.

[3] A. K. Barros and N. Ohnishi, "Pre-filtering non-stationary signals to improve blind source separation", presented at *13th International Conference On Digital Signal Processing*, vol. 2, pages 953–5, 1997.

[4] M. J. T. Alphey, D. I. Laurenson, and A. F. Murray, "The effect of signal non-stationarity on the performance of information-maximisation-based blind separation", presented at *Neural Networks for Signal Processing VIII - Proceedings of the 1998 IEEE Workshop*, pages 113–22. IEEE, 1998. To be presented at the 1998 IEEE Workshop on Neural Networks for Signal Processing.

[5] A. J. Bell and T. J. Sejnowski, "An information-maximisation approach to blind separation and blind deconvolution", *Neural Computation*, vol. 7, no. 6, pp. 1129–59, November 1995.

[6] P. Comon, "Independent component analysis, a new concept?", *Signal Processing*, vol. 36, no. 3, pp. 287–314, April 1994.

[7] H. H. Yang and S. I. Amari, "Adaptive online learning algorithms for blind separation: Maximum entropy and minimum mutual information", *Neural Computation*, vol. 9, no. 7, pp. 1457–82, October 1997.

[8] S. Amari, A. Cichocki, and H. H. Yang, "A new learning algorithm for blind signal separation", In D. S. Touretzky, M. C. Mozer, and M. E. Hasselmo, editors, *Advances in Neural Information Processing Systems*, pages 757–63, November 1996.

[9] R. Linsker, "An application of the principle of maximum information preservation to linear systems", In D. S. Touretzky, editor, *Advances in Neural Information Processing Systems*, vol. 1, pages 186–94, 1989.

Appendix B
Statistical Analyses

The ANOVA function of Genstat [119] was used for all of the statistical analyses reported in this thesis. The ANOVA tables, and other associated output, from all relevant data are included on the accompanying CD. Each experiment is contained within a unique directory, identified by the chapter in which the results were reported, and the name of the algorithm used in the experiment.

References

[1] A. Papoulis, *Probability, Random variables and Stochastic Processes*. McGraw-Hill, 2nd ed., 1984.

[2] K. Torkkola, "Blind deconvolution, information maximization and recursive filters," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 4, pp. 3301–3304, 1997.

[3] K. Torkkola, "Blind separation of convolved sources based on information maximization," in *Neural Networks for Signal Processing [1996] VI. Proceedings of the 1996 IEEE Signal Processing Society Workshop*, pp. 423–432, Sept 1996.

[4] K. Torkkola, "IIR filters for blind deconvolution using information maximization," in *NIPS'96 Workshop on Blind Signal Processing*, (Snowmass CO.), 7 Dec. 1996.

[5] K. Torkkola, "Blind separation of delayed sources based on information maximization," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, IEEE, May 1996.

[6] S. Van Gerven, D. Van Compernolle, H. Nguyen-Thi, and C. Jutten, "Blind separation of sources: A comparative study of a 2-nd and 4-th order solution," in *Signal Processing VII: Theories and Applications*, 1994.

[7] S. Van Gerven and D. Van Compernolle, "Parameter sensitivity in blind signal separation," in *Proc. Final COST 229 Workshop on Adaptive Algorithms in Communications*, (Vigo, Spain), pp. 53–57, Oct. 1994.

[8] S. Van Gerven, *Adaptive Noice Cancellation and Signal Separation with Applications to Speech Enhancement*. PhD thesis, Katholieke Universiteit Leuven, March 1996.

[9] H.-L. Nguyen Thi and C. Jutten, "Blind source separation for convolutive mixtures," *Signal Processing*, vol. 45, pp. 209–29, 1995.

[10] A. J. Bell and T. J. Sejnowski, "An information-maximisation approach to blind separation and blind deconvolution," *Neural Computation*, vol. 7, pp. 1129–59, November 1995.

[11] R. Linsker, "An application of the principle of maximum information preservation to linear systems," in *Advances in Neural Information Processing Systems*, vol. 1, pp. 186–94, 1989.

[12] L. Balmer, *Signals and Systems : An Introduction*. Prentice Hall, 1991.

[13] S. Haykin, *Adaptive Filter Theory*. Prentice Hall, second ed., 1991.

[14] R. Shaffer, "Simulated annealing." http://chem1.nrl.navy.mil/ shaffer/optsa.html.

[15] A. Farina, *Antenna-Based Signal Processing Techniques for Radar Systems*. Artech House, 1992.

[16] G. Burel and N. Rondel, "Neural networks for array processing : From DOA estimation to blind separation of sources," in *Conference Proceedings. 1993 International Conference on Systems, Man and Cybernetics. Systems Engineering in the Service of Humans*, vol. 2, pp. 601–6, October 1993.

[17] S. Choi and A. Cichocki, "Adaptive blind separation of speech signals: Cocktail party problem," in *International Conference on Speech Processing (ICSP'97)*, pp. 617–22, 1997.

[18] S. Haykin, *Neural Networks: A Comprehensive Foundation*. Prentice-Hall, Inc., 2nd ed., 1999.

[19] G. Desodt and D. Muller, "Complex independent component analysis applied to the separation of radar signals," in *Signal Processing V : Theories and Applications*, vol. I, pp. 665–8, September 1990.

[20] E. Chaumette, P. Comon, and D. Muller, "ICA technique for radiating sources estimation: application to airport surveillance," *IEE Proceedings-F*, vol. 140, pp. 395–401, December 1993.

[21] S. Li and T. J. Sejnowski, "Adaptive separation of mixed broad-band sound sources with delays by a beamforming Hérault-Jutten network," *IEEE Journal of Oceanic Engineering*, vol. 20, pp. 73–9, January 1995.

[22] K. Torkkola, "Blind separation of radio signals in fading channels," in *Advances in Neural Information Processing Systems*, vol. 10, December 1997.

[23] M. Feng and K.-D. Kammeyer, "Application of source separation algorithms for mobile communication environment," in *Proc. First International Workshop on Independent Component Analysis and Signal Separation (ICA'99)*, (Aussois, France), pp. 431–436, January 11–15, 1999.

[24] T. Ristaniemi and J. Joutsensalo, "On the performance of blind source separation in CDMA downlink," in *Proc. First International Workshop on Independent Component Analysis and Signal Separation (ICA'99)*, (Aussois, France), pp. 437–442, January 11–15, 1999.

[25] A. Prieto, B. Prieto, C. Puntonet, A. Cañas, and P. Martín-Smith, "Geometric separation of linear mixtures of sources: application to speech signals," in *Proc. First International Workshop on Independent Component Analysis and Signal Separation (ICA'99)*, (Aussois, France), pp. 295–300, January 11–15, 1999.

[26] S. Makeig, A. J. Bell, T.-P. Jung, and T. J. Sejnowski, "Independent component analysis of electroencephalographic data," in *Advances in Neural Information Processing Systems*, vol. 8, pp. 145–51, 1995.

[27] T.-Z. Jung, C. Humphries, T.-W. Lee, S. Makeig, M. J. McKeown, V. Iragui, and T. J. Sejnowski, "Removing electroencephalographic artifacts : Comparison between ICA

and PCA," in *Neural Networks for Signal Processing VIII - Proceedings of the 1998 IEEE Workshop*, (Cambridge, UK), pp. 63–72, September 1998.

[28] M. Borschbach and M. Schulte, "Performance analysis of learning rules for the blind separation of magnetoencephalography signals," in *Proc. First International Workshop on Independent Component Analysis and Signal Separation (ICA'99)*, (Aussois, France), pp. 341–346, January 11–15, 1999.

[29] A. K. Barros and N. Ohnishi, "Removal of quasi-periodic sources from physiological measurements," in *Proc. First International Workshop on Independent Component Analysis and Signal Separation (ICA'99)*, (Aussois, France), pp. 185–190, January 11–15, 1999.

[30] A. Ypma and P. Pajunen, "Rotating machine vibration analysis with second-order independent component analysis," in *Proc. First International Workshop on Independent Component Analysis and Signal Separation (ICA'99)*, (Aussois, France), pp. 37–42, January 11–15, 1999.

[31] G. Gelle, M. Colas, and G. Delaunay, "Separation of convolutive mixtures of harmonic signals with a temporal approach. Application to rotating machine monitoring," in *Proc. First International Workshop on Independent Component Analysis and Signal Separation (ICA'99)*, (Aussois, France), pp. 109–114, January 11–15, 1999.

[32] A. Back and A. Weigend, "A first application of independent component analysis to extracting structure from stock returns," *Int. Journal of Neural Systems*, vol. 8, pp. 473–484, August 1997.

[33] A. Puga and A. P. Alves, "An experiment on comparing PCA and ICA in classical transform image coding," in *Proc. First International Workshop on Independent Component Analysis and Signal Separation (ICA'99)*, (Aussois, France), pp. 105–108, January 11–15, 1999.

[34] S. Hochreiter and J. Schmidhuber, "LOCOCODE performs nonlinear ICA without knowing the number of sources," in *Proc. First International Workshop on Independent Component Analysis and Signal Separation (ICA'99)*, (Aussois, France), pp. 149–154, January 11–15, 1999.

[35] M. Girolami, "Hierarchic dichotomizing of polychotomous data—an ICA based data mining tool," in *Proc. First International Workshop on Independent Component Analysis and Signal Separation (ICA'99)*, (Aussois, France), pp. 197–202, January 11–15, 1999.

[36] A. Paraschiv-Ionescu, C. Jutten, and G. Bouvier, "Finite precision hardware implementation of source separation algorithms," in *Proc. First International Workshop on Independent Component Analysis and Signal Separation (ICA'99)*, (Aussois, France), pp. 127–132, January 11–15, 1999.

[37] C. Jutten and J. Hérault, "Analog implementation of a permanent unsupervised learning algorithm," in *Proceedings of the NATO Advanced Workshop on Neurocomputing : Algorithms, Architectures and Applications*, vol. 68, pp. 145–52, 1990.

[38] M. H. Cohen and A. G. Andreou, "Current-mode subthreshold CMOS implementation of the Hérault-Jutten autoadaptive network," *IEEE Journal of Solid-State Circuits*, vol. 27, pp. 714–27, May 1992.

[39] L. Zhang, S. Amari, and A. Cichocki, "Natural gradient approach to blind separation of over- and undercomplete mixtures," in *Proc. First International Workshop on Independent Component Analysis and Signal Separation (ICA'99)*, (Aussois, France), pp. 455–460, January 11–15, 1999.

[40] J. Karhunen, "Neural approaches to independent component analysis and source separation," in *Proc. 4th European Symposium on Artificial Neural Networks (ESANN'96)*, pp. 249–66, April 1996.

[41] J.-L. Lacoume, "A survey of source separation," in *Proc. First International Workshop on Independent Component Analysis and Signal Separation (ICA'99)*, (Aussois, France), pp. 1–6, January 11–15, 1999.

[42] S.-I. Amari, "Blind source separation — mathematical foundations," in *Brain-Like Computing and Intelligent Information Systems* (S.-I. Amari and N. Kasabov, eds.), ch. 7, Springer-Verlag, 1998.

[43] E. Oja, J. Karhunen, A. Hyvärinen, R. Vigario, and J. Hurri, "Neural independent component analysis — approaches and applications," in *Brain-Like Computing and Intelligent Information Systems* (S.-I. Amari and N. Kasabov, eds.), ch. 8, Springer-Verlag, 1998.

[44] M. Girolami, *Self-Organising Neural Networks: Independent Component Analysis and Blind Signal Separation*. Springer-Verlag, 1999.

[45] K. Torkkola, "Blind separation of audio signals: Are we there yet?," in *Proc. First International Workshop on Independent Component Analysis and Signal Separation (ICA'99)*, (Aussois, France), pp. 239–244, January 11–15, 1999.

[46] J. Anemüller and T. Gramss, "On-line blind separation of moving sound sources," in *Proc. First International Workshop on Independent Component Analysis and Signal Separation (ICA'99)*, (Aussois, France), pp. 331–334, January 11–15, 1999.

[47] Y. Naudet, M.Haritopoulos, and A. Billat, "BSS application in non-stationary mixing context," in *Proc. First International Workshop on Independent Component Analysis and Signal Separation (ICA'99)*, (Aussois, France), pp. 271–276, January 11–15, 1999.

[48] N. Parga and J.-P. Nadal, "Blind source separation with time-dependent mixtures," *Signal Processing*, vol. 80, no. 10, pp. 2187–2194, 2000.

[49] R. M. Everson and S. J. Roberts, "Adaptive non-linearities, the decorrelating manifold and non-stationary mixing for ICA," *Journal of VLSI Signal Processing-Systems for Signal, Image, and Video Technology*, 1999. (To appear).

[50] B. A. Pearlmutter and L. C. Parra, "A context-sensitive generalization of ICA," in *Proceedings of the 1996 International Conference on Neural Information Processing*, September 1996.

[51] K. Matsuoka, M. Ohya, and M. Kawamoto, "A neural net for blind separation of nonstationary signals," *Neural Networks*, vol. 8, no. 3, pp. 411–9, 1995.

[52] K. Matsuoka and M. Kawamoto, "A neural net for blind separation of nonstationary signal sources," in *1994 IEEE International Conference on Neural Networks. IEEE World Congress on Computational Intelligence.*, vol. 1, pp. 221–26, June/July 1994.

[53] A. K. Barros and N. Ohnishi, "Pre-filtering non-stationary signals to improve blind source separation," in *13th International Conference On Digital Signal Processing*, vol. 2, pp. 953–5, 1997.

[54] S. Amari, T.-P. Chen, and A. Cichocki, "Nonholonomic orthogonal learning algorithms for blind source separation." (Unpublished), 1998.

[55] E. W. Weisstein, "Eric weisstein's world of mathematics." http://mathworld.wolfram.com.

[56] C. Jutten and J. Hérault, "Blind separation of sources, Part I : An adaptive algorithm based on neuromimetic architecture," *Signal Processing*, vol. 24, pp. 1–10, July 1991.

[57] A. Hyvärinen and E. Oja, "Simple neuron models for independent component analysis," *Int. Journal of Neural Systems*, vol. 7, no. 6, pp. 671–87, 1996.

[58] S. Amari, A. Cichocki, and H. H. Yang, "A new learning algorithm for blind signal separation," in *Advances in Neural Information Processing Systems*, pp. 757–63, November 1996.

[59] T.-W. Lee, A. J. Bell, and R. Orglmeister, "Blind source separation of real world signals," in *1997 IEEE International Conference on Neural Networks (ICNN'97)*, vol. 4, pp. 2129–34, IEEE, June 1997.

[60] J.-F. Cardoso and B. Laheld, "Equivariant adaptive source separation," *IEEE Transactions on Signal Processing*, vol. 44, no. 12, pp. 3017–30, 1996.

[61] A. Cichocki and Moszczyński, "New learning algorithm for blind separation of sources," *Electronic Letters*, vol. 28, pp. 1986–7, October 1992.

[62] A. Cichocki, R. Unbehauen, and E. Rummert, "Robust learning algorithm for blind separation of signals," *Electronic Letters*, vol. 30, pp. 1386–7, August 1994.

[63] P. Comon, "Contrasts for multichannel blind deconvolution," *IEEE Signal Processing Letters*, vol. 3, pp. 209–11, July 1996.

[64] D. Yellin and E. Weinstein, "Criteria for multichannel signal separation," *IEEE Transactions on Signal Processing*, vol. 42, pp. 2158–68, August 1994.

[65] E. Moreau and B. Stoll, "An iterative block procedure for the optimization of constrained contrast functions," in *Proc. First International Workshop on Independent Component Analysis and Signal Separation (ICA'99)*, (Aussois, France), pp. 59–64, January 11–15, 1999.

[66] M. Girolami and C. Fyfe, "Higher order cumulant maximisation using nonlinear Hebbian and anti-Hebbian learning for adaptive blind separation of source signals," in *Proceedings of the Third International Workshop on Image and in Computational Intelligence (IWISPO'96)*, pp. 141–4, 1996.

[67] J. L. Lacoume and P. Ruiz, "Separation of independent sources from correlated inputs," *IEEE Transactions on Signal Processing*, vol. 40, pp. 3074–8, December 1992.

[68] P. Comon, "Independent component analysis, a new concept?," *Signal Processing*, vol. 36, pp. 287–314, April 1994.

[69] N. Delfosse and P. Loubaton, "Adaptive blind separation of independent sources : A deflation approach," *Signal Processing*, vol. 45, pp. 59–83, 1995.

[70] C. Fyfe and M. Girolami, "Tracking independent sources," in *Proceedings of the Second ICSC International Symposium on Soft Computing (SOCO'97)*, September 1997.

[71] J.-F. Cardoso and A. Souloumiac, "Blind beamforming for non Gaussian signals," *IEE Proceedings-F*, vol. 140, pp. 362–370, Dec. 1993.

[72] D. J. C. MacKay, "Maximum likelihood and covariant algorithms for independent component analysis." http://www.inference.phy.cam.ac.uk/mackay/README.html, August 1996.

[73] J.-F. Cardoso, "Infomax and maximum likelihood for blind source separation," *IEEE Signal Processing Letters*, vol. 4, pp. 112–3, 1997.

[74] *Proc. First International Workshop on Independent Component Analysis and Signal Separation (ICA'99)*, (Aussois, France), January 11–15 1999.

[75] J. Hérault, C. Jutten, and B. Ans, "Détection de grandeurs primatives dans un message composite par une architecture de calcul neuromimétique un apprentissage non supervisé," in *Proc. Xème colloque GRETSI*, (Nice, France), pp. 1017–22, May 1985.

[76] P. Comon, C. Jutten, and J. Hérault, "Blind separation of sources, Part II : Problems statement," *Signal Processing*, vol. 24, pp. 11–20, July 1991.

[77] E. Sorouchyari, "Blind separation of sources, Part III : Stability analysis," *Signal Processing*, vol. 24, pp. 21–9, July 1991.

[78] Y. Deville, "A unified stability analysis of the Hérault-Jutten source separation neural network," *Signal Processing*, vol. 51, pp. 229–33, 1996.

[79] P. Comon, "Statistical approach to the Jutten-Hérault algorithm," in *Proceedings of the NATO Advanced Workshop on Neurocomputing : Algorithms, Architectures and Applications*, vol. 68 of *Computer & System Sciences*, pp. 81–8, 1990.

[80] T.-P. Jung, S. Makeig, M. Westerfield, J. Townsend, E. Courchesne, and T. Sejnowski, "Independent component analysis of single-trial event-related potentials," in *Proc. First International Workshop on Independent Component Analysis and Signal Separation (ICA'99)*, (Aussois, France), pp. 173–178, January 11–15, 1999.

[81] A. Hyvärinen and E. Oja, "One-unit learning rules for independent component analysis," in *Advances in Neural Information Processing Systems*, vol. 9, pp. 480–6, 1997.

[82] A. Hyvärinen and E. Oja, "A fast fixed-point algorithm for independent component analysis," *Neural Computation*, vol. 9, pp. 1483–92, 1997.

[83] J. Karhunen and P. Pajunen, "Blind source separation and tracking using nonlinear PCA criterion: A least-squares approach," in *Proceedings of the 1997 IEEE International Conference on Neural Networks, (ICNN'97)*, (Houston, Texas), pp. 2147–52, June 9–12 1997.

[84] E. Oja, "Nonlinear PCA criterion and maximum likelihood in independent component analysis," in *Proc. First International Workshop on Independent Component Analysis and Signal Separation (ICA'99)*, (Aussois, France), pp. 143–148, January 11–15, 1999.

[85] M. Girolami and C. Fyfe, "Negentropy and kurtosis as projection pursuit indices provide generalised ICA algorithms," in *Advances in Neural Information Processing Systems (NIPS96) Blind Signal Separation Workshop*, December 1996.

[86] M. Girolami and C. Fyfe, "Blind separation of sources using exploratory projection pursuit networks," in *Proceedings of International Conference on the Engineering Applications of Neural Networks*, vol. 1, pp. 249–52, 1996.

[87] M. Kawamoto, A. Barros, A. Mansour, K. Matsuoka, and N. Ohnishi, "Blind separation for convolutive mixtures of non-stationary signals," in *Fifth International Conference on Neural Information Processing*, (Kitakyushu, Japan), pp. 743–746, 21-23 October 1998.

[88] C. Fyfe and A. Cichocki, "Hierarchical and parallel models for non-stationary independent component analysis," *Recent Advances in Soft Computing*, 1998.

[89] J. C. Platt and F. Faggin, "Networks for separation of sources that are superimposed and delayed," in *Advances in Neural Information Processing Systems* (S. J. Hanson, J. E. Moody, and R. P. Lippmann, eds.), vol. 4, pp. 730–737, Morgan Kaufman, 1991.

[90] T.-W. Lee, A. J. Bell, and R. H. Lambert, "Blind separation of delayed and convolved sources," in *Advances in Neural Information Processing Systems*, vol. 9, pp. 758–64, 1997.

[91] R. H. Lambert, *Multichannel Blind Deconvolution: Source Separation with Multipath*. PhD thesis, University of Southern California, 1996.

[92] T.-W. Lee, A. Ziehe, R. Orglmeister, and T. J. Sejnowski, "Combining time-delayed decorrelation and ICA: Towards solving the cocktail party problem," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, (Seattle), pp. 1249–1252, May 1998.

[93] L. Molgedey and H. Schuster, "Separation of a mixture of independent signals using time delayed correlations," *Physical Review Letters*, vol. 72, pp. 3634–3647, June 1994.

[94] S. Choi, S. Amari, A. Cichocki, and R. Liu, "Natural gradient learning with a nonholonomic constraint for blind deconvolution of multiple channels," in *First International Workshop on Independent Component Analysis and Signal Separation*, (Aussois, France), pp. 371–376, 1999.

[95] S. Amari, S. C. Douglas, A. Cichocki, and H. H. Yang, "Multichannel blind deconvolution and equalization using the natural gradient," in *First IEEE Signal Processing Workshop on Signal Processing Advances in Wireless Communications*, April 1997.

[96] S. Van Gerven and D. Van Compernolle, "Signal separation by symmetric adaptive decorrelation: Stability, convergence, and uniqueness," *IEEE Transactions on Signal Processing*, vol. 43, pp. 1602–1612, July 1995.

[97] B. S. Krongold and D. L. Jones, "Blind source separation of nonstationary convolutively mixed signals," in *Proceedings of the 10th IEEE Workshop on Statistical Signal and Array Processing*, (Pocono Manor, PA), pp. 53–57, 2000.

[98] A. Koutras, E. Dermatas, and G. Kokkinakis, "Simultaneous speech recognition in noisy reverberant environments," in *CSCC 2000*, (Athens, Greece), 2000.

[99] A. Koutras, E. Dermatas, and G. Kokkinakis, "Simultaneous speech recognition in noisy reverberant environments," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2000)*, (Istanbul, Turkey), 2000.

[100] C. Mejuto, A. Dapena, and L. Castedo, "Frequency-domain infomax for blind separation of convolutive mixtures," in *Proceedings of the Second International Workshop on Independent Component Analysis and Blind Signal Separation (ICA2000)* (P. Pajunen and J. Karhunen, eds.), (Helsinki, Finland), pp. 315–320, June 2000.

[101] S. Ikeda and N. Murata, "A method of ICA in time-frequency domain," in *Proc. First International Workshop on Independent Component Analysis and Signal Separation (ICA'99)*, (Aussois, France), pp. 365–376, January 11–15, 1999.

[102] R. H. Lambert and C. L. Nikias, *Unsupervised Adaptive Filtering, Volume 1*, ch. 9 — Blind Deconvolution of Multipath Mixtures. John Wiley & Sons, 2000.

[103] P. Smaragdis, "Efficient blind separation of convolved sound mixtures," in *IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics*, 1997.

[104] P. Smaragdis, "Blind separation of convolved mixtures in the frequency domain," in *International Workshop on Independence & Artificial Neural Networks*, (University of La Laguna, Tenerife, Spain), 9-10 February 1998. Revised version also in Neurocomputing 22 (1998) pp. 21–34.

[105] L. Parra and C. Spence, "Convolutive blind separation of non-stationary sources," *IEEE Transactions on Speech and Audio Processing*, vol. 8, pp. 320–327, May 2000.

[106] L. Parra and C. Spence, "On-line blind source separation of non-stationary signals," *Journal of VLSI Signal Processing*, vol. 26, no. 1/2, 2000.

[107] C. Fancourt and L. Parra, "The coherence function in blind source separation of convolutive mixtures of non-stationary signals," in *Proceedings of the 2001 IEEE Signal Processing Society Workshop, Neural Networks for Signal Processing XI*, pp. 303–312, 2001.

[108] D. Schobben and P. Sommen, "A new blind signal separation algorithm based on second order statistics," in *Proceedings of the IASTED International Conference on Signal and Image Processing*, (Las Vegas, USA), pp. 564–569, October 1998.

[109] D. Schobben and P. Sommen, "A new algorithm for joint blind signal separation and acoustic echo canceling," in *Proceedings of the Fifth International Symposium on Signal Processing and Its Applications (ISSPA)*, vol. 2, pp. 889–892, 1999.

[110] F. J. Owens, *Signal Processing of Speech*. MacMillan, 1993.

[111] L. Rabiner and B.-H. Juang, *Fundamentals of Speech Recognition*. Prentice Hall, 1993.

[112] H. H. Yang and S.-I. Amari, "Adaptive online learning algorithms for blind separation: Maximum entropy and minimum mutual information," *Neural Computation*, vol. 9, pp. 1457–82, October 1997.

[113] B. Arons, "Techniques, perception, and applications of time-compressed speech," in *Proceedings of 1992 Conference, American Voice I/O Society*, pp. 169–77, September 1992.

[114] J. Murphy, *Resource Allocation in ATM Networks*. PhD thesis, Dublin City University, 1996.

[115] ATM Forum, "Voice Networking in the WAN." http://www.atmforum.com/atmforum/library/vtoa.html.

[116] S. Policker and A. B. Geva, "Non-stationary signal analysis using temporal clustering," in *Neural Networks for Signal Processing VIII - Proceedings of the 1998 IEEE Workshop*, (Cambridge, UK), pp. 304–12, September 1998.

[117] D. Brook and R. J. Wynne, *Signal Processing : principles and applications*. Edward Arnold, 1988.

[118] R. H. Lambert, "Difficulty measures and figures of merit for source separation," in *Proc. First International Workshop on Independent Component Analysis and Signal Separation (ICA'99)*, (Aussois, France), pp. 133–138, January 11–15, 1999.

[119] Genstat 5 Committee (1987), *Genstat 5 Reference Manual*. Oxford: Clarendon Press, 1987.

[120] J. McNicol , Personal communication.

[121] M. O. Pun and Y. Hirai, "A simple variable step algorithm for blind source separation (BSS)," in *Neural Networks for Signal Processing VIII - Proceedings of the 1998 IEEE Workshop*, (Cambridge, UK), pp. 304–12, September 1998.

[122] A. Westner and V. M. Bove, Jr., "Blind separation of real world audio signals using overdetermined mixtures," in *Proc. First International Workshop on Independent Component Analysis and Signal Separation (ICA'99)*, (Aussois, France), pp. 251–256, January 11–15, 1999.

[123] D. Schobben and P. Sommen, "On the indeterminacies of convolutive blind signal separation based on second order statistics," in *Proceedings of the Fifth International Symposium on Signal Processing and Its Applications (ISSPA)*, vol. 1, pp. 215–218, 1999.

[124] The MathWorks, Inc., "Matlab v12.1."

[125] P. Smaragdis, "Synthetic bench." http://sound.media.mit.edu/ica-bench/code/.

[126] D. Schobben, K. Torkkola, and P. Smaragdis, "Evaluation of blind signal separation methods," in *Proc. First International Workshop on Independent Component Analysis and Signal Separation (ICA'99)*, (Aussois, France), pp. 261–266, January 11–15, 1999.

[127] M. J. T. Alphey, D. I. Laurenson, and A. F. Murray, "The effect of signal non-stationarity on the performance of information-maximisation-based blind separation," in *Neural Networks for Signal Processing VIII - Proceedings of the 1998 IEEE Workshop*, (Cambridge, UK), pp. 113–22, September 1998.

[128] M. J. T. Alphey, D. I. Laurenson, and A. F. Murray, "Improvements in the on-line performance of information-maximisation-based blind signal separation," in *Proceedings of the First International Workshop on Independent Component Analysis and Blind Signal Separation (ICA'99)*, pp. 49–54, January 1999.

[129] M. J. T. Alphey, D. I. Laurenson, and A. F. Murray, "The effect of signal non-stationarity on the performance of information-maximisation-based blind separation," *PhDEE — The Postgraduate Journal of the Department of Electronics and Electrical Engineering*, pp. 59–65, April 1998.